

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-114616

(43)Date of publication of application : 02.05.1995

(51)Int.Cl.

G06K 9/00

G06F 19/00

(21)Application number : 05-262100

(71)Applicant : HITACHI LTD

(22)Date of filing : 20.10.1993

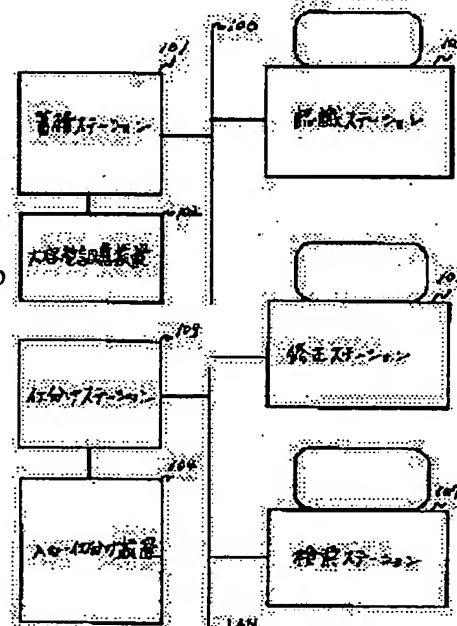
(72)Inventor : SHIMA YOSHIHIRO  
HATANO HIDEKAZU  
MARUKAWA KATSUMI  
KOGA MASASHI  
NAKAJIMA KAZUKI  
KADOTA AKIZO  
KURINO KIYOMICHI  
SUGIMOTO TAKEYUKI

## (54) SLIP DOCUMENT INFORMATION SYSTEM

(57)Abstract:

**PURPOSE:** To examine the adequacy of entry contents by discriminating the kinds of the formats of plural slip documents, and automatically setting a read area and performing character recognition.

**CONSTITUTION:** The slip document information system is constituted by connecting a sorting station 103, a storage station 101, a recognition station 105, a correction station 106, and a retrieval station 107 to a network 100. Consequently, plural slip documents can be read. Further, the adequacy of entry contents is examined and the entry contents of a slip can automatically be inspected. Paper documents can automatically be sorted to reduce the sorting operation.



## LEGAL STATUS

[Date of request for examination]

28.03.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of

$I_b = [(La+Lb+1)/2]$ 、 $I_c = [(La+Lc+1)/2]$ 、 $I_d = [(La+Lb+Lc+Ld+2)/4]$

を用いる丸め込み方法であり、前記マイナスの丸め込み方法は、

$I_b = [(La+Lb)/2]$ 、 $I_c = [(La+Lc)/2]$ 、 $I_d = [(La+Lb+Lc+Ld+1)/4]$

を用いる丸め込み方法であることを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、フレーム間予測を行い、輝度または色の強度が量子化された数値として表される動画像符号化および復号化方法、ならびに動画像の符号化装置および復号化装置に関するものである。

【0002】

【従来の技術】動画像の高効率符号化において、時間的に近接するフレーム間の類似性を活用するフレーム間予測（動き補償）は情報圧縮に大きな効果を示すことが知られている。現在の画像符号化技術の主流となっている動き補償方式は、動画像符号化方式の国際標準である

$$P(x,y) = R(x+u_i, y+v_i) \quad (x,y) \in B_i, 0 \leq i < N \quad \dots \text{ (数1)}$$

【0005】で表される。ただし、画像はN個のブロックに分割されるとして、 $B_i$ は画像のi番目のブロックに含まれる画素、 $(u_i, v_i)$ はi番目のブロックの動きベクトルを表している。

【0006】 $u_i$ と $v_i$ の値が整数ではないときには、参照画像において実際には画素が存在しない点の強度値を

$$R(x+\frac{p}{d}, y+\frac{q}{d}) = ((d-q)((d-p)R(x,y) + pR(x+1,y)) + q((d-p)R(x,y+1) + pR(x+1,y+1))) // d^2 \quad \dots \text{ (数2)}$$

【0008】で表される。ただし「//」は除算の一種で、通常の除算（実数演算による除算）の結果を近隣の整数に丸め込むことを特徴としている。

【0009】図1にH. 263の符号化器の構成例100を示す。H. 263は、符号化方式として、ブロックマッチングとDCT（離散コサイン変換）を組み合わせたハイブリッド符号化方式（フレーム間／フレーム内適応符号化方式）を採用している。

【0010】減算器102は入力画像（現フレームの原画像）101とフレーム間／フレーム内符号化切り換えスイッチ119の出力画像113（後述）との差を計算し、誤差画像103を出力する。この誤差画像は、DCT変換器104でDCT係数に変換された後に量子化器105で量子化され、量子化DCT係数106となる。この量子化DCT計数は伝送情報として通信路に出力されると同時に、符号化器内でもフレーム間予測画像を合成するために使用される。

【0011】以下に予測画像合成の手順を説明する。上述の量子化DCT係数106は、逆量子化器108と逆

H. 263、MPEG1、MPEG2に採用されている半画素精度のブロックマッチングである。この方式では、符号化しようとする画像を多数のブロックに分割し、ブロックごとにその動きベクトルを水平・垂直方向に隣接画素間距離の半分の長さを最小単位として求める。

【0003】この処理を数式を用いて表現すると以下のようなになる。符号化しようとするフレーム（現フレーム）の予測画像Pの座標(x, y)におけるサンプル値（輝度または色差の強度のサンプル値）をP(x, y)、参照画像R（Pと時間的に近接しており、既に符号化が完了しているフレームの復号画像）の座標(x, y)におけるサンプル値をR(x, y)とする。また、xとyは整数であるとして、PとRでは座標値が整数である点に画素が存在すると仮定する。また、画素のサンプル値は負ではない整数として量子化されているとする。このとき、PとRの関係は、

【0004】

【数1】

求めることが必要となる。この際の処理としては、周辺4画素を用いた共1次内挿が使われることが多い。この内挿方式を数式で記述すると、dを正の整数、 $0 \leq p$ 、 $q < d$ として、 $R(x+p/d, y+q/d)$ は、

【0007】

【数2】

DCT変換器109を経て復号誤差画像110（受信側で再生される誤差画像と同じ画像）となる。これに、加算器111においてフレーム間／フレーム内符号化切り換えスイッチ119の出力画像113（後述）が加えられ、現フレームの復号画像112（受信側で再生される現フレームの復号画像と同じ画像）を得る。この画像は一旦フレームメモリ114に蓄えられ、1フレーム分の時間だけ遅延される。したがって、現時点では、フレームメモリ114は前フレームの復号画像115を出力している。この前フレームの復号画像と現フレームの入力画像101がブロックマッチング部116に入力され、ブロックマッチングの処理が行われる。

【0012】ブロックマッチングでは、画像を複数のブロックに分割し、各ブロックごとに現フレームの原画像に最も似た部分を前フレームの復号画像から取り出すことにより、現フレームの予測画像117が合成される。このときに、各ブロックが前フレームと現フレームの間でどれだけ移動したかを検出する処理（動き推定処理）を行う必要がある。動き推定処理によって検出された各

rejection]

[Date of requesting appeal against examiner's decision of  
rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 特 許 公 報 (B 2)

(11) 特許番号

特許第3092613号  
(P3092613)

(45) 発行日 平成12年9月25日 (2000.9.25)

(24) 登録日 平成12年7月28日 (2000.7.28)

(51) Int.Cl.

識別記号

F I

H 0 4 N 7/32

H 0 4 N 7/137

Z

G 0 6 F 7/38

G 0 6 F 7/38

H 0 3 M 7/36

H 0 3 M 7/36

請求項の数 3 (全 16 頁)

(21) 出願番号 特願平11-30968  
(62) 分割の表示 特願平10-160205の分割  
(22) 出願日 平成10年6月9日 (1998.6.9)  
  
(65) 公開番号 特開平11-317955  
(43) 公開日 平成11年11月16日 (1999.11.16)  
審査請求日 平成11年3月30日 (1999.3.30)  
(31) 優先権主張番号 特願平9-150656  
(32) 優先日 平成9年6月9日 (1997.6.9)  
(33) 優先権主張国 日本 (J P)

(73) 特許権者 000005108  
株式会社日立製作所  
東京都千代田区神田駿河台四丁目6番地  
(72) 発明者 中屋 雄一郎  
東京都国分寺市東恋ヶ窪1丁目280番地  
株式会社日立製作所 中央研究所内  
(72) 発明者 棚寝 義人  
東京都国分寺市東恋ヶ窪1丁目280番地  
株式会社日立製作所 中央研究所内  
(74) 代理人 100075096  
弁理士 作田 康夫

審査官 畑中 高行

最終頁に続く

(54) 【発明の名称】 記録媒体

(57) 【特許請求の範囲】

【請求項1】 動画像の符号化情報が複数のフレームに関する情報として記録された記録媒体において、前記フレームに関する情報は、動き補償によって求められた現フレームの予測画像と前記現フレームの入力画像との差分に関する情報と、前記動き補償によって求められた動きベクトルに関する情報と、前記動き補償がプラスの丸め込みによる動き補償かマイナスの丸め込みによる動き補償かを識別する情報とが記録されていることを特徴とする記録媒体。

【請求項2】 動画像に関する情報が符号化ビットストリームとして記録された記録媒体において、前記動画像は動き補償を用いて符号化され、前記符号化ビットストリームは、前記動き補償において画素の強度値の内挿に用いられた丸め込み方法がプラスの丸め込み方法或いはマ

イナスの丸め込み方法のいずれかであることを識別する情報を有することを特徴とする記録媒体。

【請求項3】 請求項2に記載の記録媒体において、前記プラスの丸め込み方法は、前記動画像を構成する画像の強度Laを有する第1の画素と、前記第1の画素と水平方向に隣接し強度Lbを有する第2の画素と、前記第1の画素と垂直方向に隣接し強度Lcを有する第3の画素と、前記第2の画素と垂直方向に隣接し、かつ前記第3の画素と水平方向に隣接し強度Ldを有する第4の画素とから、画素の存在しない前記第1の画素と前記第2の画素の間接点における強度Ibと、前記第1の画素と前記第3の画素の間接点における強度Icと、前記第1と第2と第3と第4の画素とに囲まれ前記第1、第2、第3、および第4の画素から等距離にある点における強度Idとを求める際に、

## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

## CLAIMS

---

### [Claim(s)]

[Claim 1] The cut-form document information system characterized by having a means to discriminate the modality of format of the concerned specification, a means to read according to the concerned modality of format and to extract a field, a means to set up a reading field and to perform character recognition, and a means to authorize the validity of the written content on the basis of the data of the concerned recognition result, in the cut-form document information system of the medical treatment fee bill of particulars which performs a picture image input, a store, reference, and content check at least.

[Claim 2] In claim 1 publication, a means to discriminate the modality of format of the aforementioned specification A means to key the modality of specification, and a means to store the modality of format as dictionary data beforehand, The cut-form document information system which has a means to collate a means to extract format data from the inputted specification document, and the format data extracted with the concerned dictionary data, and is characterized by limiting collating with dictionary data according to the modality of the concerned specification which keyed.

[Claim 3] The cut-form document information system characterized by having a means to discriminate the modality of cut-form document and to perform character recognition in the cut-form document information system of a paper document which performs picture image input, store, and reference at least, a means to \*\*\*\* a reference number in a paper document by the recognition result, and a means to classify a paper document according to the modality of the concerned document.

[Claim 4] The cut-form document information system characterized by having a means to discriminate the modality of cut-form document and to perform character recognition in the cut-form document information system of a paper document which performs picture image input, store, and reference at least, and a means to print a bar code on the front face of a paper document according to the concerned recognition result.

[Claim 5] In the document information system of a paper document which performs picture image input, store, display, and reference at least A means to set up two or more fields of a document picture image as a viewing area, and a means to read in the concerned two or more fields, and to extract a field, The document information system characterized by having a means to display simultaneously the picture image and recognition result of the concerned viewing area on a screen, and the means which reads when correcting a recognition result through a screen, and carries out the enlarged display of the picture image of a field.

[Claim 6] In the cut-form document information system of a paper document which performs a picture image input, a store, reference, and content check at least A means to discriminate the modality of format of a cut-form document, and a means to extract the field where the content of \*\*\*\* was indicated according to the concerned modality of format, The cut-form document information system characterized by having a means to carry out the mask of the picture image of the concerned \*\*\*\* field, and to display it, and a means by which the level of the password given to the user performs cancel of the concerned mask and a setup.

[Claim 7] The cut-form document information system characterized by having a means to extract the inputted cut-form document picture image to a closing line, a means to register the closing line of a standard cut-form document, a means to choose the extracted closing line and a similar closing line from registration data, and a means to display simultaneously the concerned selected registration closing line and an extraction closing line on a screen, in the cut-form document information system which performs the picture image input of a paper document, and registering a cut-form document.

[Claim 8] In the cut-form document information system of a paper document which performs a picture image input, a store, reference, and content check at least A means to extract the reading field of a cut-form document, and a means to perform character recognition in a reading field, A means to judge whether there is any numeric data of a

regards,

Shingo Kawase  
of Marco Polo Club

Visit our Web site @: <http://myweb.to/marco>

recognition result within limits set up beforehand, The cut-form document information system characterized by having a means to accumulate a cut-form document according to the concerned judgment result, choosing the cut-form document picture image judged that is in concerned within the limits, and accumulating and searching only the concerned picture image.

[Claim 9] The document information system characterized by having the input station which inputs the surface picture image of a paper document, the recognition station which discriminates the modality of inputted picture image, the classification station which \*\*\*\*s a paper document according to an identification result, and the store station which saves a picture image in the document information system which classifies a paper document through a network, and transmitting image data, the concerned modality data, and classification control information via a network.

[Claim 10] The classification station characterized by generating beeping and suspending classification processing if the picture image of a paper document is inputted, it has the stacker which stores the hopper which is the classification station which recognizes the concerned paper document picture image and discriminates a modality, and sets up a paper document, and the paper document after classification and a stacker fills.

[Claim 11] The document information system characterized by having a means to store a document picture image and its reference number, a means to ask for the bar code for arrangement by specifying the reference number of a document, a means to read the bar code of the front face of a paper document, and a means to choose the document to which the specified bar code was given, in the document information system of a document which performs a picture image input, a store, reference, and content check at least.

---

[Translation done.]



**From:** ReedSmith-IPO3.Hazel <ReedSmith-IPO3.Hazel@c-net3.crl.hitachi.co.jp>  
**To:** Hiroshi-ipo3.Asamitsu@c-net3.crl.hitachi.co.jp <Hiroshi-ipo3.Asamitsu@c-net3.crl.hitachi.co.jp>  
**Date:** Sunday, November 04, 2001 11:13 PM  
**Subject:** Re:310002267 point of novelty

---

Based upon the meeting, the major prior art reference is extremely close to the invention. Please confer with the inventor to see how long will it takes him to find out an appropriate point of novelty to overcome the prior art reference. Although any physical dimensional limitation of the inspection pads is argumenably a design choice, as long as the results/benefits are not anticipated by the prior art, we may still argue that arrangement is patentable in a crowded art as the semiconductor prakaging design.

My revision is pending for the point of novelty, and I prefer to close the case by 11/30 since I have to finish all cases before I leave.

Jennifer Teng  
ext 2878

## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

## DETAILED DESCRIPTION

---

### [Detailed Description of the Invention]

[0001]

[Field of the Invention] this invention relates to the cut-form document information system which performs reading of two or more kinds of cut-forms (for example, medical treatment fee bill of particulars etc.) from which a format and format are different, classification, archive, reference, etc.

[0002]

[Description of the Prior Art] Although the format of a document specifies the order of a written item and an item, it does not specify the absolute position coordinate and a dimension. The format of a document not only specifies the order of a written item and an item, but has specified the absolute position coordinate and the dimension. With the conventional optical-character-recognition equipment, it had set up beforehand, having used format of a document as format data, and the paper document with which the absolute position coordinate of each item of a document and the dimension are specified was able to be read. However, the document with which only the format is specified, for example, a statute document etc., was not able to be read.

[0003] The mark posting field which directs the item which should be beforehand inputted into a cut-form is prepared, and the method which therefore reads the predetermined field on a cut-form to read the mark entered in the concerned field is proposed as the document information system which reads the conventional cut-form is stated to JP,56-11573,A (Japanese-Patent-Application-No. 87338 [ 54 to ] July 10, Showa 54 (1979) application, artificer:Masayuki Kagawa, and applicant:Fujitsu, Inc.). However, in the conventional method, since the mark posting field was prepared, the area on the space of the concerned posting field to occupy increased, and there was a problem that an input item could not be increased. Moreover, since it is necessary to newly prepare a mark posting field, there is a problem that the various existing cut-forms cannot be read as they are. Furthermore, by the cut-form document the format is decided to be by law etc., since the format was specified in the eye of the law, preparing a mark posting field like before does not accept, but it had the problem that the concerned format document without a mark posting field could not be read.

[0004] Moreover, a feature parameter is extracted from the field of the document containing the bar graph which is the photograph and binary graphic picture image which are a halftone picture image, and the image-file method which relates with the picture image of a field and is memorized is proposed as the file method of the conventional document picture image is stated to JP,59-14070,A (Japanese-Patent-Application-No. 123460 [ 57 to ] July 15, Showa 57 (1982) application, artificer:Koichi Ejiri, applicant:Ricoh Co., Ltd.). However, discriminating the document which consists of a frame which consists of a ruled line and a ruled line like a cut-form only in performing identification of a photograph and a bar graph being described, and a character string is not described by the conventional method. Moreover, it is not taken into consideration about the file method which carries out the mask of carrying out character recognition of the character string of a cut-form, and the field of a document according to the grade of \*\*\*\*.

[0005] Moreover, the modality of document is automatically discriminated to a fixed form input-statement document, and the filing method which gives the same classification code as the already stored document is stated to JP,61-75477,A (Japanese Patent Application No. 196616 [ 59 to ], September 21, Showa 59 (1984) application, artificer:Yasuaki Nakano, Hiromichi Fujisawa, Masaaki Fujinawa, applicant:Hitachi, Ltd.). However, in this conventional method, it being main storing a picture image, setting up a reading field, and performing character recognition or authorizing the validity of the written content on the basis of the data of the concerned recognition result are not taken into consideration. For this reason, the written content cannot be read according to the modality of document, and the validity of the content cannot be authorized.

[0006]



[Problem(s) to be Solved by the Invention] it had been difficult to look for the paper document which is the original itself, although technique can display and print a desired document picture image from the image file which giving a reference number or a bar code to the inputted paper document itself is not indicated, but was accumulated since each above-mentioned \*\*, and the work which looks for the original was work which needs many helps Moreover, classifying the paper document itself is not taken into consideration, but a sort of the original, arrangement, and receipt had taken many helps.

[0007] moreover, since each above-mentioned \*\*, with technique, even when it is not taken into consideration about a \*\*\*\* item when displaying and printing the accumulated document picture image, for example, the important information is indicated by the document on the private content and the business, the picture image of a reference result is displayed on a screen as it is, or the entire display has been forbidden for every picture image unit – it is that either and the subregion of a document was not able to be seen according to \*\*\*\* level.

[0008] moreover, since each above-mentioned \*\*, with technique, it is a premise to accumulate all the inputted paper documents, and selecting and accumulating has not been taken into consideration according to the content For this reason, a lot of documents will be accumulated and there was a problem exceed the capacity of storage, in a small-scale file system.

[0009] It is in the 1st purpose of this invention discriminating the modality of format of two or more cut-form documents, setting up a reading field automatically, performing character recognition, and authorizing the validity of the written content.

[0010] It is in the 2nd purpose of this invention printing a reference number or a bar code in a paper document, and classifying a paper document by the recognition result.

[0011] It is in the 3rd purpose of this invention extracting the field where the content of \*\*\*\* was indicated from a document picture image, carrying out the mask of the picture image of a \*\*\*\* field, and performing display of a reference result, and printing.

[0012] It is in the 4th purpose of this invention judging the existence of the need for a store automatically, and saving a storage region according to the written content of a document.

[0013] The 5th purpose of this invention is to offer the document information system which performs a recognition, classification, and a store through a network.

[0014]

[Means for Solving the Problem] In order to attain the 1st above-mentioned purpose, the modality of format is discriminated from a document picture image, the written content is read, and it judges whether the conditions which the reading result has registered for every written item are satisfied.

[0015] Moreover, in order to attain the 2nd above-mentioned purpose, a classification device and a print station are controlled by the recognition result.

[0016] Moreover, in order to attain the 3rd above-mentioned purpose, the subregion the item was indicated to be is automatically extracted from a document picture image, and the password for an access is given to a \*\*\*\* field.

[0017] In order to attain the 4th above-mentioned purpose, the written content, for example, mark, is read and only the document with mark larger than a constant value accumulates the surface picture image.

[0018] In order to attain the 5th above-mentioned purpose, while a picture image is compressed and transmitted to a network, a recognition result is transmitted and the classification station and store station which were connected to the network by the data are controlled.

[0019]

[Function] The surface picture image is inputted to two or more kinds of cut-forms, and the modality of format of a cut-form document is discriminated. And it reads according to the modality of format, a field is extracted, and character recognition is performed. The validity of the written content is authorized on the basis of the data of a recognition result according to the criteria set up beforehand. Furthermore, by the recognition result, \*\*\*\* or a bar code is printed for a reference number in a paper document, and a paper document is classified according to the modality of document. Moreover, the field where the content of \*\*\*\* was indicated according to the modality of document is extracted, and the mask of the picture image of a \*\*\*\* field is carried out, and it enables it to display it.

[0020]

[Example] Hereafter, this invention is explained in detail using a drawing. Drawing 1 is drawing explaining the configuration of the cut-form information system which is one example of this invention.

[0021] Input / classification station 103, the store station 101, the recognition station 105, the correction station 106, and the reference station 107 are connected to the network 100. Input / classification station 103 consists of a

From: marco\_polo@geocities.co.jp <marco\_polo@geocities.co.jp>  
 To: marco\_polo@egroups.co.jp <marco\_polo@egroups.co.jp>  
 Date: Thursday, November 01, 2001 2:34 AM  
 Subject: [ML] #21 Dumpling party, Nov 18 (registration)

---

(English is following.)

□\$B!Z<j:n\$j;q-;R%Q!<%F%#%\$%Y%s%H\$N;22CEPO?3+;O!□(B

□\$B\$\_J\$5\$^!"\$3\$s\$K\$A\$O!#!!%^^k%3%]!<%m6f3ZIt\$G\$9!#□(B

□\$BK!F|h\$j!%"^%k%3%]!<%m6f3ZIt\$G\$O!h□(B21□\$B2s%\$%Y%s%H\$X□(B  
 □\$B\$N;22CEPO?r3+;O\$7\$^\$9!#!!\$^\$:O%a!<%j%s%0%j%9%H\$K□(B  
 □\$BEPO?5\$!\$F\$\$\$kJ)\$N\$\_\$rBP>]H\$7\$^\$9!#!!□(B

□\$BF|;~!□(B 11□\$B7n□(B18□\$BF|!JF|!K□(B □\$B8a8e□(B  
 □\$BFbMF!□(B □\$B<j:n\$j;q-;R%Q!<%F%#□(B  
 □\$B>|~j!□(B □\$B@EDC+6hFb\$N8x6&!\@\_□(B  
 □\$BHqMQ!□(B 1000□\$B1\_□(B □\$B!JM=Dj!K!!<c43A)8:\$9\$k3\$H\$, \$"\$j\$^\$9□(B  
 □\$BDj0w!!!!□(B15□\$BL>!!!J=i2sDy\$a@Z\$jJ,!K□(B  
 □\$B=89g!?!7=11X□(B 11:40□\$B!\$^\$?O!!□(B  
 □\$BE!5^@\$EDC+@~!&@\$EDC+1X□(B 12:30

\* □\$B=89g>|~j\$J!\$N>\\$7\$\$\$pJs\$O;22CEPO?\$r\$5\$!\$?J)□(B  
 □\$B!!\$K\$\*CN\$!\$;\$7\$^\$9!#□(B

□\$BK!G/:G8e\$O%^^k%3%]!<%m6f3ZIt\$G\$b?M5\$\$,\$" \$kNAM}\$rBj□(B  
 □\$B:~\$H\$7\$^\$9!#!!NAM)\$OKh2s!VF1\$83x\$NHS\$r?)\$&!W\$H\$\$\$&8@MU□(B  
 □\$B\$b\$"\$k\$H\$\*\$j!";22C<T\$, \$^\$H\$^\$C\$?N!\$J7005\$\$\$K\$J\$j\$^\$9!#□(B

□\$B%^^k%3%]!<%m6f3ZIt\$K;22C\$r4uK>\$5\$!\$kJ)\$O!"3hF0J)?K!&□(B  
 □\$BFbMF\$rCN\$C\$F\$\$\$?@\$/\$?\$a\$K!"%[!<%`%Z!<%8>e\$N!V%^^k%3□(B  
 □\$B%)!<%m6f3ZIt\$K\$D\$\$\$F!W\$H!V3hF0FbMF!W\$O;vA0\$K\$\*F!\$\_\$/□(B  
 □\$B\$@\$\$\$\$^\$9\$h\$&\$\*4j\$\$\$7\$^\$9!#□(B  
 □\$B!!□(B  
 □\$B\$\*?=\$79~\$\_\$K:]\$7\$^\$7\$F\$O!"□(B

□\$B!z□(B □\$B\$\*L>A0□(B □\$B!J4A;z%U%k%M!<%`!K□(B  
 □\$B!z□(B □\$B\$\*L>A0□(B □\$B!J% m!<%^;z%U%k%M!<%`!K□(B  
 □\$B!z□(B □\$B=89g>|~j\$O\$!\$A\$!\$+!!!J?7=11X□(B or □\$B@EDC+1X!K□(B

□\$B\$r5-:\\$N\$&\$(!"\$3\$N%a!<%k\$KJV?.\$7\$F\$/\$@5\$\$\$!#□(B

□\$BEPO?\$O!";22C<TK!M\$,9T\$C\$F\$/\$@5\$\$\$!#!!3+;O\$h\$j#3F|□(B  
 □\$B4V\$O!"%a!<%j%s%0%j%9%H9XFI<T\$N\$\_EPO?\$,\$G\$-\$^\$9!#□(B

□\$BJ#?t?M?i\$G\$N;22C\$r4uK>\$5\$!\$k>I9g\$b!"EPO?\$O8DJL\$K□(B  
 □\$B9T\$C\$F\$/\$@5\$\$\$!#!!\$\_J\$5\$^\$N\$46(NO\$r\$\*4j\$C\$W\$7\$^\$9!#□(B

□\$B86B'\$H\$7\$F@hCe=g\$K<u\$1IU\$1\$^\$9\$,!"DjOw\$K\$C#\$7\$?>I□(B  
 □\$B9g\$O!V7+\$j>e\$, \$jBT\$A!W\$H\$J\$K\$3\$H\$, \$"\$j\$^\$9\$N\$G\$4N;>5□(B  
 □\$B\$/\$@5\$\$\$!#□(B

11/15/2001

computer. The input and the assortment equipment 104 which performs picture image input of input / classification cut-form document and classification of a cut-form document are connected to the concerned station 103. The hopper which sets up a document, the stacker which stores the document after classification, and its control section are in the concerned equipment, and optical system, such as an optical image pick-up element for inputting a picture image and a lighting system, possesses further. Here, the data compression of the inputted picture image is carried out, and image data is sent out to a network 100 through the concerned station 103. Moreover, a document is classified, receiving a recognition result from a network 100 and using this as a classification control signal. It consists of a computer, the large capacity storage 102 which consists of an optical disk or an array disk is connected, and the store station 101 performs the data of a recognition result, and the inputted store of a picture image. At the store station 101, the document which should be accumulated based on a recognition result can be sorted out. Moreover, to the subregion the content of \*\*\*\* is indicated to be, an access information is given and it accumulates to the concerned storage 102. At the recognition station 105, sheet identification processing in which the modality of cut-form document is searched for is performed to the picture image sent through a network 100, and the character string indicated by the item is recognized. At the correction station 106, correction of a recognition result is interactively performed using a screen. Here, a reading field is expanded, it displays on a screen, and correction is made easy. The cut-form picture image accumulated at the store station 101 can be searched through a network, the reference station 107 can be searched for the cut-form of a request of the content indicated by the item at a key, and it can display on a screen. At this time, the access information is given to the subregion the content of \*\*\*\* is indicated to be, and the mask of the concerned \*\*\*\* item can be carried out with the level of an access of a user, and it can display.

[0022] Drawing 2 is drawing explaining the processing flow of a cut-form document. As an example of a cut-form document, the processing flow of the medical treatment fee bill of particulars (medical statement) is shown. Data are inputted about the main items indicated by the medical statement, and a computer performs authentication processing of capability record. In this processing flow, a medical statement is first set as the hopper for an input in step 200. Subsequently, classification processing of a medical statement is performed at step 201. Here, it classifies by discriminating the format as a modality of medical statement in the visitor part medical statement 210 and the hospitalization part medical statement 211. An initial information is inputted at step 202. Here, distinction of a self-prefecture and another prefecture is inputted as an initial information as a local attribute for processing a medical statement in each prefecture unit. Moreover, the initial value of the office name and medical-examination partition which process a medical statement, and a reference number is registered. In addition, about the hospitalization part medical statement 211, a reference number is \*\*\*\*ed at step 208. Subsequently, reading is performed at step 203 and step 204. Step 203 is batch reading processing in which batch processing of a cut-form is performed, and it prints a reference number while it stores a part for non-\*\* (\*\*\*\*\*\*) in another stacker at the normal part row of a recognition result. Step 204 is processing of correction reading serially, and corrects a recognition result interactively through a correction screen to the medical statement which it rejected as a result of the recognition. The data for the norm recognized at step 203 and the data which became normal to be corrected at step 204 are stored in a floppy disk etc. Moreover, the medical statement which became \*\*\*\*\* at step 203 will be stored in another stacker. In addition, a part for \*\*\*\*\* of a hospitalization part medical statement and step 203 is collected, and punch work is performed at step 209. On the basis of the content given in a medical statement of medical examination, as shown in step 207, machine processing of capability record authentication processing etc. is performed. Here, only the medical statement with which large sum mark are indicated is chosen, and it stores in an optical disk at step 205. In authentication processing of capability record of step 207, reference 206 of the medical statement stored in the optical disk at step 205 is performed, a desired medical statement is displayed on a screen, and capability record is checked.

[0023] Drawing 3 is the processing flow of the system which performs the picture image input of a cut-form document, a store, reference, and content check. The surface picture image of a cut-form document is inputted at step 300, and sheet identification processing in which the modality of format of a cut-form is discriminated at step 301 is performed. The detailed processing flow of the concerned sheet identification is mentioned later in drawing 4. Subsequently, in step 302, it reads according to the concerned modality of format, a field is extracted, and a reading field is set up. The detailed processing flow of the concerned reading field setting step 302 is mentioned later in drawing 9. In step 303, it reads, character recognition is performed in a field, and the validity of the written content is authorized on the basis of the data of the concerned recognition result at step 304. The detailed processing flow of the concerned step 303 is mentioned later in drawing 10.

[0024] Drawing 4 is drawing showing the detailed processing flow of sheet identification processing of step 301.

From: ReedSmith-IPO3.Hazel <ReedSmith-IPO3.Hazel@c-net3.crl.hitachi.co.jp>  
To: Hideji Fujii <Hideji-ipo3.Fujii@c-net3.crl.hitachi.co.jp>  
Cc: uchida@harl.hitachi.co.jp <uchida@harl.hitachi.co.jp>  
Date: Wednesday, November 14, 2001 11:27 PM  
Subject: Re: 310100082

---

Dear Mr. Fujii,

I'll forward you the physical file with my revision shortly. As you will notice, I added new claims 10-19, and marked up all the intended changes in pencil for you to review which will later be entered by Ms. Yuko Suzuki.

Please review the revision to see whether you and the inventor agree. Please provide your feedback by the noon of 11/29 such that I can finalize the this application while my memory is still fresh.

Jennifer Teng  
ext 2878

11/15/2001

Here, a frame and a character row are extracted from the inputted picture image, and the matching with the dictionary for sheet identification 430 beforehand registered from the standard document is performed. First, at step 400, a minute noise is removed from an input picture image. Subsequently, a ruled line is extracted using the run which is the segment which a black pixel follows at step 401, and the angle of inclination of the concerned ruled line is computed at step 402. Only the concerned angle of inclination rotates a picture image at step 403, and rotation correction is performed. And a ruled line is again extracted at step 404. Drawing 6 is explanatory drawing of the process which takes out a level ruled line and a perpendicular ruled line from a picture image by this processing [ a series of ], and computes the starting point coordinate and a terminal-point coordinate. It is the result of 601 extracting the level ruled lines 602, 603, and 604 to the cut-form picture image 600. Moreover, 610 is the result of extracting the perpendicular ruled lines 611-616. At step 405, the character row which is relation of a character is extracted in rectangular type, and the coordinate of the angle is searched for. That a document is level and in order to correct a position gap of a perpendicular direction, it asks for the reference point of a document at step 406. At step 407, the coordinate of the intersection which is equivalent to the angle of a frame from a level ruled line and a perpendicular ruled line is computed. Since each ruled line can be expressed by the straight-line formula, it can search for a linear intersection coordinate easily. At step 408, the normalization of size is performed to the intersection coordinate searched for. Furthermore, at step 409, a frame is generated on the basis of an intersection. Drawing 7 is drawing explaining the process which asks for an intersection from a level ruled line and a perpendicular ruled line, and subsequently generates a frame. The intersections 710-723 are computed on the basis of the level ruled lines 602-604 and the perpendicular ruled lines 611-616. And an intersection is followed one by one and frames 701, 702, -706 are generated. A frame is matched at step 410 of drawing 4. At this time, the matching with the frame of two or more canonicals stored in the dictionary for sheet identification 430 is performed. The frame position and the position of a character row are registered beforehand, and first, the flow of the registration inputs a standard cut-form picture image into the concerned dictionary 430 at step 420, and, subsequently specifies the position of the frame and character row which should be inputted at step 421 to be it from a screen. At step 422, correction of a position and rotation is performed and it registers as a dictionary. Furthermore, a character row matching is performed at step 411. At this time, the matching with the character row of two or more canonicals stored in the dictionary for sheet identification 430 is performed. The detail of a character row matching is later mentioned in drawing 8. According to the result searched for by the frame matching 410 described above and the character row matching 411, the sheet modality number which a cut-form most similar among standard cut-forms has is determined at step 412. Furthermore, the generated frame position and the position of a character row are outputted at step 413. Moreover, a sheet kind number is outputted at step 414.

[0025] Drawing 5 shows the recognition system of coordinates which have a position gap, rotation, and expansion and contraction. The frame ruled line shown in the cut-form sheet 500 by 506 exists. Recognition system of coordinates are expressed by the recognition system of coordinates (Xm, Ym) 512 which have expansion and contraction by the recognition system of coordinates (Xr, Yr) 511 without expansion and contraction. These system of coordinates are observation system of coordinates which consist of a sensor and optical system. Moreover, the system of coordinates for format data generation 513 (Xf, Yf) are the system of coordinates on the basis of the edge of a sheet, and are the system of coordinates showing the dimension of life. The amount (dx, dy) of gaps of the zero 510 of recognition system of coordinates and the zero 509 of the system of coordinates for format data generation and angle-of-rotation theta can express the target point 507 in the sheet expressed by the recognition system of coordinates without expansion and contraction (xr, yr).

[0026]

[Equation 1]  $x_r = dx + x_f$ ,  $\cos\theta + y_f$ ,  $\sin\theta + y_r = dy - x_f$ ,  $\sin\theta + y_f$ , and  $\cos\theta - (x_f, y_f)$  are the coordinates of the target point 507 expressed with the system of coordinates for format data generation here Furthermore, the target point 507 (xm, ym) in the sheet expressed by the recognition system of coordinates with expansion and contraction is degrees of shrinkage mx and my [0027]

[Equation 2] It is [0028] when  $x_m = mx$  and  $x_{ym} = my$ , and  $y_r$ .

[Equation 3]

$x_m = m(dx + x_f, \cos\theta + y_f, \text{and} \sin\theta)$

$y_m = m(dy - x_f, \sin\theta + y_f, \text{and} \cos\theta)$

It can express. In reading in a sheet, according to an upper formula, format data are changed, a picture image is accessed, and character recognition is performed. Moreover, the value of the recognition system of coordinates of the reference point for position doubling 508 is expressed with the recognition system of coordinates (xr0, yr0) without expansion and contraction, and the recognition system of coordinates (xm0, ym0) with expansion and



**From:** ReedSmith-IPO3.Hazel <ReedSmith-IPO3.Hazel@c-net3.crl.hitachi.co.jp>  
**To:** k-kojima@crl.hitachi.co.jp <k-kojima@crl.hitachi.co.jp>  
**Cc:** Hideji Fujii <Hideji-ipo3.Fujii@c-net3.crl.hitachi.co.jp>  
**Date:** Thursday, November 15, 2001 3:43 AM  
**Subject:** 310100400

---

Dr. Kojima,

The application is ready after I clarify two questions with you. Please reply as soon as possible.

1. Should "inorganic" be added to the support in the claims such that the claim will not be overly board?
2. Is it necessary to change "an organic solvent" into "organic solvent"? It is not necessary if we adopt the second definition as follows.

**SOLVENT Chemistry.** 1st. A substance in which another substance is dissolved, forming a solution. 2nd. A substance, usually a liquid, capable of dissolving another substance.

Jennifer Teng  
ext 2878

contraction. Moreover, the value of the system of coordinates for format data generation is expressed with (xf0, yf0). The amount (dx, dy) of position gaps is [0029].

[Equation 4] It can express with  $dx = (1/mx)$ ,  $xm0 - xf0$ ,  $\cos\theta - yf0$ ,  $\sin\theta dy = (1/my)$ ,  $ym0 + xf0$ ,  $\sin\theta - yf0$ , and  $\cos\theta$ . Angle-of-rotation  $\theta$  is detected on the basis of the inclination of a ruled line. Subsequently, although asked for degrees of shrinkage  $mx$  and  $my$ , this detects the breadth of the linea-nigra frame of most outside, and a dip, and computes them as compared with the breadth and the dip to which format data correspond. Furthermore, it asks for the position ( $xm0$ ,  $ym0$ ) in the recognition system of coordinates of the reference point for position doubling, and the amount (dx, dy) of position gaps is calculated as compared with the position ( $xf0$ ,  $yf0$ ) (it expresses by the system of coordinates for format data generation) of the concerned point beforehand set as format data. By sheet identification, the coordinate ( $xm$ ,  $ym$ ) of the target point detected in the recognition system of coordinates with expansion and contraction is changed into the system of coordinates for format data generation ( $xf$ ,  $yf$ ), and this is compared with the coordinate data of the format of the dictionary for sheet identification.

[0030] The method expressing the relative physical relationship of a character row is shown in drawing 29. Now, the physical relationship of eight orientation is expressed to the character row 2900 to observe. When orientation is expressed in north, south, east and west, there is a character row 2911 in the orientation 2901 of north of the character row 2900 to observe. Moreover, in the orientation 2905 of south, there has a character row 2913 in the orientation 2903 of an east, and a character row 2917 has a character row 2915 in the orientation 2907 of west, respectively. furthermore – the orientation 2902 of a northeast of the character row 2900 to observe – a character row 2912 – in the orientation 2908 of a northwest, there has a character row 2914 in the orientation 2904 of southeast, and a character row 2916 has a character row 2918 in the orientation 2906 of southwest, respectively. Thus, the character row which has the physical relationship of eight orientation is expressed to the character row to observe.

[0031] Drawing 8 is drawing explaining mode of processing of a character row matching of step 411. The example of the subregion of the character row of a cut-form document is shown in 810, 811, and 812. the character row extracted in the character row field 810 – 801 and 802 – although shown 803, such relative physical relationship is up-and-down north-south relations. When the character row to observe is now set to 801, there is a character row 802 in the orientation 820 of south to a character row 801. Moreover, similarly there is a character row 803 in the orientation 821 of south to a character row 802. On the other hand, in the example of the character row field 811, the relative physical relationship of character rows 804, 805, and 806 is an east-and-west relation on either side. To a character row 804, there is a character row 805 in the orientation 822 of an east, and it has 806 in the orientation 823 of an east to a character row 805. Furthermore, in the example of the character row field 812, it is in the physical relationship of the orientation of slant, and there is a character row 808 in the orientation 824 of a northeast to a character row 807. Moreover, similarly there is a character row 809 in the orientation 825 of a northeast to a character row 808. In a character row matching of step 411, the relative physical relationship of such a character row is beforehand registered into the dictionary for sheet identification 430, and it judges whether it agrees with the relative physical relationship of the character row extracted from the input picture image.

[0032] Drawing 9 is the detailed processing flow of reading field setting processing shown at step 302 of drawing 3. The sheet kind number obtained at step 301 of previous sheet identification is inputted at step 900. Subsequently, it reads at step 901, and moves to processing for every field, and the format parameter of corresponding reading is read from the format parameter 910 registered for every sheet modality at step 902. Here, as a format parameter, the position of a character kind, the number of characters, and a reading field possesses. At step 903, a reading field is computed on the basis of the amount of position gaps from the reference point position which was read from the format parameter and for which was read and it asked at a field position and step 406.

[0033] Drawing 10 is the detailed processing flow of character recognition processing in a reading field shown at step 303. It reads at step 1000 and processing to steps 1001-1004 is performed for every field. At step 1001, it reads and the character row in a field is extracted. At step 1001, the unnecessary character currently printed in the reading field is removed beforehand. And the character row extracted at step 1003 is separated per single character. At step 1004, character recognition for every single character is performed. This character recognition is easily realizable using the conventional technique.

[0034] Drawing 11 is drawing showing the processing flow of the cut-form document information system which reads a cut-form document, and chooses and accumulates a cut-form document on the basis of the concerned reading result. 1101 or less-step processing is repeated until a document is exhausted at step 1100. The picture image input of a cut-form document is performed at step 1101. Subsequently, the same sheet identification processing as step 301 is performed at step 1102. Subsequently, similarly a reading field is extracted at step 1103, it

Inbox for jtinjp@yahoo.com

Yahoo! - My Yahoo! Options - Sign Out - Help



Mail Addresses Calendar Notepad

**Lose 10 Pounds By Dec. 17th** eDiets  
Height  ft  in Weight  lbs

[Prev](#) | [Next](#) | [Inbox](#)

**From:** Yasunari-ipo3.Imaki@c-net3.crl.hitachi.co.jp | [Block Address](#) | [Add to Address Book](#)  
**To:** jteng@reedsmith.com  
**Date:** Mon, 12 Nov 2001 12:09:50 +0900  
**Subject:** Please correct the specification #310100317

> -----  
> From:  
Yasunari-ipo3.Imaki@c-net3.crl.hitachi.co.jp [SMTP:YASUNARI-IPO3.IMAKI@C-NET3.CRL.HITACHI.CO.J  
> Sent: Sunday, November 11, 2001 10:09:50 PM  
> To: Teng, Jennifer  
> Subject: Please correct the specification #310100317  
> Sensitivity: Personal  
> Auto forwarded by a Rule  
>

Dear Ms. Jennifer Teng  
from Y. Imaki @IPO3, Hitachi, Ltd.

Please correct "Ohm-sha, denshi-tsuushin-jouhou-gakkai-hen" to  
"published by Ohm-sha, and edited by The Institute of Electronics,  
Information and Communication Engineers --- IEICE" in the Specification page 1  
of #310100317.

----- Begin of forwarded message -----

TO: Yasunari-ipo3.Imaki  
w@?;\$ Masahiro-crl.Kaminaga  
?;\$ ????a??;\$01/11/12 11:49  
-;\$ ?.

?\_?  
?G?;

?????;?G?;?

??;??p???

The Institute of Electronics, Information and Communication  
Engineers --- IEICE

??;G?;

----- End of forwarded message -----

Click a to send an instant message to an online friend

= Online, = Offline

[Prev](#) | [Next](#) | [Inbox](#)

reads at step 1104, and character recognition is performed in a field. It judges whether at step 1105, there is any numeric data of a recognition result within limits set up beforehand, and when it is within set-up limits, the concerned picture image is accumulated to storage, such as an optical disk, at step 1106. When there is nothing within set-up limits, the concerned picture image is not accumulated to storage. A series of above-mentioned processing is performed until a document is exhausted at step 1100. It is necessary to accumulate not all the paper documents inputted by this processing, and they can be selected and accumulated according to the written content. For this reason, in a small-scale file system, it is effective in the ability to save the capacity of storage.

[0035] In the document information system which performs the picture image input of a cut-form document, a store, reference, and content check, according to a user's access level, the mask of the picture image of a \*\*\*\* field is carried out, and it is displayed. Drawing 12 shows the processing flow which registers a \*\*\*\* field from a cut-form document picture image. A picture image is inputted at step 1200 and sheet identification processing 301 is performed at step 1201. And a \*\*\*\* field is extracted at step 1202. Here, the position of the field extracted with the sheet kind at step 1201 of previous sheet identification is outputted, and it judges whether it is a \*\*\*\* field for every field. At this time, whether it is a \*\*\*\* field shall read the dictionary for sheet identification 430 registered beforehand. The level of \*\*\*\* possesses for every field shown in the concerned dictionary 430 by the frame. The coordinate of a \*\*\*\* field is stored in storage with image data at step 1203.

[0036] Drawing 13 is the flow of processing which carries out the mask of the picture image of a \*\*\*\* field by a user's access level, and is displayed to a cut-form picture image. First, a user's password is inputted at step 1300. And a user's access level is set up at step 1301 using the user access-level table 1600 shown in drawing 16. Drawing 16 is explanatory drawing of a user access table. To the user name 1601 and the password 1602, the access level 1603 is stored and the access level corresponding to the concerned user is set up. Subsequently, processing of steps 1303 and 1304 is repeated until a field is exhausted at step 1302. At step 1303, the access level set up for every field is read from the field access table 1500. Drawing 15 shows a field access table. The access level 1502 is given every identification number 1501 by which the concerned table 1500 was given to the field to the standard cut-form document. And it judges whether a user's access level is stronger than the access level of a field at step 1304. When a user's access level is strong, the mask of the concerned field is canceled at step 1305. On the other hand, on the contrary, when [ that ] the access level of a field is stronger than a user's access level, a mask is set as the concerned field at step 1306. A series of above processing is repeatedly performed until a field is exhausted at step 1302. And a cut-form picture image is expressed as step 1307. Here, the picture image which performed display processing for every field, and did not display the partial picture image by which the mask is set as the field as it was, for example, was smeared away by the black picture image is displayed. The partial picture image of which the mask of a field is canceled is displayed as it is. In this example, when displaying and printing the accumulated document picture image, the access level is given about the \*\*\*\* item, for example, when the important information is indicated on the content private in a document, and the business, the subregion of a document can be seen according to \*\*\*\* level, and it is effective in the ability to maintain secrecy.

[0037] Drawing 14 is an example of a display of a cut-form document picture image. In the display screen 1400, the fields 1401, 1402, 1403, 1404, 1415, 1405, 1406, 1407, 1408, and 1409 corresponding to each item of a cut-form are set up. According to a user's access level, in the oblique-line fractions 1410, 1411, 1412, 1413, and 1414, the mask is set up for \*\*\*\* and the concerned field will be smeared away black at this time.

[0038] Drawing 17 shows the processing flow which prints a bar code in the concerned document by the reading result of a cut-form document. A document picture image is inputted at step 1700. Sheet identification which discriminates the modality of cut-form document at step 1701 is performed to the inputted picture image, and the field in a sheet is read by character recognition at step 1702. And according to the concerned recognition result, a bar code is printed on the front face of a paper document at step 1703.

[0039] Drawing 18 shows the processing flow which looks for a document by the bar code. If the reference number of a document to look is inputted at step 1800, the bar code which corresponds to a reference number at step 1801 will be set up. And processing of steps 1803, 1804, and 1805 is repeated until a document is exhausted at step 1802. First, in step 1803, the ejection of a cut-form document is performed and the bar code currently printed by the paper front face is read at step 1804. Subsequently, in step 1805, the read bar code judges whether it is in agreement with the bar code set up at step 1801. It stores in a selection stacker at step 1806 noting that it is a desired cut-form document, if in agreement. Processing is repeated until it stores in the usual stacker in step 1807 and a document is exhausted at step 1802, if not in agreement at step 1805. Since the bar code is printed in the original by this example, it is effective in the ability to look for the paper document which is the original itself easily using a bar code it not only to display and to print a desired document picture image from the accumulated image file, but.

composition of the photosensitive polymer layer.

>

>Jennifer Tng

>

>

Drawing 19 is a processing flow which classifies by discriminating the modality of cut-form document. A cut-form document picture image is inputted at step 1900, and the dictionary for sheet identification 1910 performs sheet identification at step 1901. Subsequently, reading in a sheet is performed at step 1902. At this time, the format parameter 1911 and the dictionary for character recognition 1912 which have been registered beforehand are used. At step 1903, a cut-form document is classified according to the modality of sheet which it is as a result of sheet identification. By this example, the cut-form document currently conventionally performed by the help can be automatically classified now, and the rating by the help can be reduced.

[0040] Drawing 20 is drawing explaining the motion control of input / classification station connected to the network, and a recognition station. Input / classification station 2002 and the recognition station 2001 are connected to the network 2000. A control of processing sends activation of a whole system to input / classification station 2002 from the recognition station 2001 first. Input / classification station 2002 and this activation command are received, and the cut-form document set as the hopper inputs a picture image with the scanner of an input unit 2004. The data compression of the picture image is carried out in the picture image processor in an input unit 2004, and it is transmitted to the recognition station 2001 through a network 2000. It is elongated and the transmitted picture image performs sheet identification and reading. A recognition result is transmitted to input / classification station 2002 through a network 2000, and chooses the stacker of an assortment equipment 2005. The cut-form document with which the hopper is set up is lost, and the above-mentioned processing is performed until a hopper becomes empty. In addition, when a stacker fills with on the way, beeping is generated and it halts. If a hopper becomes empty, an end command will be transmitted to the recognition station 2001 from input / classification station 2002, and processing of a whole system will be completed.

[0041] Drawing 21 is the flow of the cut-form document processing system which classifies by discriminating the modality of cut-form document, performing character recognition, and \*\*\*\*ing a reference number in a paper document by the recognition result. The picture image of a cut-form document is inputted at step 2100. Sheet identification which discriminates the modality of cut-form at step 2101 is performed. Reading in a sheet is performed at step 2102, and a reference number is \*\*\*\*ed at step 2103 by the concerned recognition result. At step 2104, it classifies according to the modality or written content of a cut-form, and stores in a predetermined stacker.

[0042] Drawing 22 is an example of the cut-form document which \*\*\*\*ed the reference number. The reference number shown by 2201 is \*\*\*\*ed in the cut-form document 2200. Since the position of the edge of a document is detected, the concerned reference number can be \*\*\*\*ed in the coordinate position set up beforehand. Moreover, the ink color for \*\*\*\* can be switched to a specification color.

[0043] Drawing 23 is drawing showing the configuration of an input and classification. The cut-form document 2320 is set as the hopper 2300. While 2304 is a computer and whole input and classification are controlled. Communications control with a network 2321 is performed. The hopper control section 2310 controls the \*\*\*\*\* device section 2315, and sends paper to the sensor section 2313 one by one. The sensor-control section 2311 digitizes a picture signal, and sends it to a computer 2304. 2314 is a \*\*\*\* device and prints the reference number sent by the \*\*\*\* control section 2312 from the computer 2304. The classification device sections 2316, 2317, and 2318 are controlled by the command of a computer 2304 to store the sent cut-form document in a predetermined stacker. The selected cut-form document is stored in stackers 2301, 2302, and 2303 one by one. In each stacker, the sensors 2307, 2308, and 2309 which detect \*\*\*\* of a document possess, and a stacker full signal is sent to a computer 2304 by the full detection sensor-control section 2306. 2305 is the beeping occurrence section, if \*\*\*\* of a stacker is detected, beeping will be generated and a computer 2304 will suspend processing.

[0044] Drawing 24 is an example of the transition diagram of the screen of a cut-form document information system. The main screen 2401 at the time of a system startup is shown in drawing 25. The main screen 2500 possesses the input prompt 2502 of the selections 2501 and selections. Two kinds of inputs can be switched with the input technique specification screen 2402, one of them is performed on the scanner input screen 2403, and another is performed on the file input screen 2404. 2405 is an image display screen, can input a picture image by the specified input technique, and can carry out display authentication of the input picture image on the concerned screen. From the concerned screen, it can return to the main screen 2401, for example, transits on a format / format identification screen 2406. Here, as shown in drawing 26, the extracted frame 2601 can be displayed on Screen 2600. Moreover, as shown in Screen 2602, the extracted frame 2603 and the frame 2604 provided as a dictionary can also be piled up and displayed. 2408 is a recognition result screen and is correctable by displaying a document picture image and a recognition result simultaneously. Drawing 27 is drawing explaining a recognition result screen. The up image display area 2701 and the lower image display area 2702 are expanded and displayed to the cut-form document 2700. In the recognition result screen 2710, the up picture image is expanded and displayed on

wherein the itemset providing means provides an itemset  $X$  containing at least two affirmative terms and applies a logic AND operation on the itemset  $X$  and a negative item  $\sim a$  to generate a candidate itemset  $X \sim a$ , before any support count or support is calculated,

wherein the sorting means also sorts out frequently used itemsets from a plurality of the candidate itemsets by ensuring each of which support is more than the minimum value of support, the system further comprising:

means for calculating an upper bound of confidence of the association rule by calculating a support count of each one of the frequently used itemsets ("Supp( $X$ )"), a support count of  $a$  ("Supp( $a$ )"), and a support count of  $X$  ("Supp( $X$ )"), for selecting a smaller value in  $\{\text{Supp}(X), \text{Supp}(a)\}$ , and for selecting a bigger value in  $\{\text{the smaller value}, \text{Supp}(X)\}$ , and

wherein the sorting means also sort out more frequently used itemsets by ensuring each of which upper bound of confidence is less than the minimum value of confidence and the association rule contains at least one of the more frequently used itemsets.

2711. Moreover, the lower picture image is expanded and displayed on 2712. A recognition result is interactively correctable by displaying on the fraction of 2713. Drawing 28 is drawing explaining other examples of a recognition result screen. In this example, the up picture image 2801 is displayed on the 1st screen 2800, using a recognition result screen as two screens, and the recognition result is displayed on the fraction of 2802. Moreover, the lower picture image 2804 is displayed, similarly a recognition result is displayed on 2805, and it corrects to the 2nd screen 2803 interactively.

[0045] Drawing 30 is the detailed processing flow of validity certification of step 304, and is an example for a medical statement. Certification of sex and the content of a claim is performed at step 3000. There are some from which the medical action differs by sex, and it authorizes whether the content of a claim, such as a check and an operation, corresponds by sex. Furthermore, certification of age and the content of a claim is performed at step 3001. Here, the content of a pharmacy or an operation has a limit according to age, and it authorizes whether the content of a claim is suitable in the concerned age. Furthermore, certification of an age and cut-form format is performed at step 3002. Here, it authorizes whether it agrees with the age when the format of the cut-form for which the format of a cut-form is beforehand specified by the neonate, infants, the old man, etc., and it asked by sheet identification according to age is indicated. Moreover, the correspondence with the name registered with the insurance certificate number at step 3003 is judged. Furthermore, in steps 3004 and 3005, certification of medical-examination real days and a totaling point is performed.

[0046]

[Effect of the Invention] The modality of format is discriminated from a document picture image, the written content is read, and the reading result is made to judge whether the conditions registered for every written item are satisfied, and can read two or more cut-form documents. Moreover, a reading field is set up automatically, character recognition is performed, the validity of the written content is authorized, and the written content of a cut-form can be checked automatically. It is made to control a classification device and a print station by the recognition result, and a reference number or a bar code can be printed in a paper document by the recognition result, a paper document can be classified automatically, and it is effective in reducing classification work. It enables it to give the password for an access to a \*\*\*\* field, the mask of the picture image of a \*\*\*\* field can be carried out, display of a reference result and printing can be performed, and it is effective in raising the reliability of a system. The written content, for example, mark, is read, and only the document with mark larger than a constant value is made to accumulate the surface picture image, and it is effective in the ability to save a storage region.

---

[Translation done.]



10. A system for mining at least one association rule based upon data in a database containing attributes, comprising:

means for defining each of the attributes with a respective attribute value as an item, and defining the database as a plurality sets of items;

means for defining an item appearing in one of the sets as an affirmative item in the set, and defining an item absent from the set as a negative item in the set; and

means for mining the association rule by applying at least one logic AND operation on an assumption of the association rule containing at least one negative item and at least one affirmative item so as to obtain a conclusion of the association rule containing at least one affirmative item.

11. The system for mining at least one association rule according to claim 10, further comprising:

means for applying at least one logic AND operation on at least two items to generate an itemset;

means for providing a plurality sets of itemsets to be included in the database;

means for defining a support count of one item as a number of times the item appearing in different said sets of items and a support count of one itemset as a number of times the itemset appearing in different said sets of itemsets;

means for calculating a support of one item or itemset by dividing a support count of the item or the itemset in the database with a total number of the item or the itemset in the database;

means for calculating a confidence of one item or itemset by dividing a support count of the item or the itemset in (the assumption AND the conclusion) with a support count of the item or the itemset in the assumption; and

means for calculating a confidence increment for each association rule containing at least one negative item by omitting the negative item in said each association rule to calculate a confidence of the association rule without the negative item so as to obtain an evaluation value, and dividing the confidence of the association rule with the evaluation value.

12. The system for mining at least one association rule according to claim 11, further comprising:

means for inputting a minimum value of support, a minimum value of confidence, and a minimum value of confidence increment of the association rule,

wherein the mining means mining the association rule by sorting out each itemset contained in the assumption of the association rule satisfying the minimum value of support, the minimum value of confidence, and the minimum value of confidence increment.

13. → The system for mining at least one association rule according to claim 12,

## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

DESCRIPTION OF DRAWINGS

---

## [Brief Description of the Drawings]

- [Drawing 1] It is drawing explaining the configuration of a cut-form information system.
- [Drawing 2] It is drawing explaining the processing flow of a cut-form document.
- [Drawing 3] It is drawing showing the processing flow of the system which performs the picture image input of a cut-form document, a store, reference, and content check.
- [Drawing 4] It is drawing showing the detailed processing flow of sheet identification processing.
- [Drawing 5] It is drawing explaining the recognition system of coordinates which have a position gap, rotation, and expansion and contraction.
- [Drawing 6] It is explanatory drawing of the process which computes a ruled line coordinate.
- [Drawing 7] It is drawing explaining the processing process of ruled line intersection calculation and frame generation.
- [Drawing 8] It is drawing explaining mode of processing of a character row matching.
- [Drawing 9] It is explanatory drawing of the detailed processing flow of reading field setting processing.
- [Drawing 10] It is explanatory drawing of the detailed processing flow of character recognition processing in a reading field.
- [Drawing 11] It is explanatory drawing of the cut-form document information system which performs a selection store of a cut-form document.
- [Drawing 12] It is explanatory drawing of the processing flow which registers a \*\*\*\* field from a cut-form document picture image.
- [Drawing 13] It is explanatory drawing of the processing flow which carries out the mask of the picture image of a \*\*\*\* field by a user's access level, and is displayed.
- [Drawing 14] It is drawing explaining an example of a display of a cut-form document picture image.
- [Drawing 15] It is explanatory drawing of a field access table.
- [Drawing 16] It is explanatory drawing of a user access-level table.
- [Drawing 17] It is explanatory drawing of the processing flow which prints a bar code in a document using a recognition result.
- [Drawing 18] It is explanatory drawing of the processing flow which looks for a document by the bar code.
- [Drawing 19] It is explanatory drawing of the processing flow which classifies by discriminating the modality of cut-form document.
- [Drawing 20] It is explanatory drawing of the motion control of a station connected to the network.
- [Drawing 21] It is explanatory drawing of the cut-form document processing system flow which classifies by \*\*\*\*ing a reference number by the recognition result.
- [Drawing 22] It is drawing showing an example of the cut-form document which \*\*\*\*ed the reference number.
- [Drawing 23] It is drawing showing the configuration of an input and classification.
- [Drawing 24] It is the transition diagram of the screen of a cut-form document information system.
- [Drawing 25] It is drawing showing a main screen.
- [Drawing 26] It is drawing showing the display screen of a frame.
- [Drawing 27] It is drawing explaining a recognition result screen.
- [Drawing 28] It is drawing explaining other examples of a recognition result screen.
- [Drawing 29] It is drawing explaining the relative physical relationship of a character row.
- [Drawing 30] It is the detailed processing flow of validity certification.

## [Description of Notations]

100 [ -- A recognition station, 103 / -- Input / classification station, 301 / -- A sheet identification processing step,

304 / -- A validity certification step, 410 / -- A frame matching step, 411 / -- A character row matching step, 910 / -- A format parameter, 1306 / -- A mask setup of a field 1703 / -- A bar-code printing step, 1804 / -- Bar-code reading, 2103 / -- Reference number \*\*\*\*, 2104 / -- Classification processing. ] -- A network, 101 -- A store station

---

[Translation done.]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-114616

(43)公開日 平成7年(1995)5月2日

(51)Int.Cl.<sup>4</sup>

識別記号

片内整理番号

F I

技術表示箇所

G 0 6 K 9/00

K 8623-5L

G 0 6 F 19/00

8724-5L

G 0 6 F 15/ 22

審査請求 未請求 請求項の数11 O L (全 19 頁)

(21)出願番号 特願平5-262100

(22)出願日 平成5年(1993)10月20日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 嶋 好博

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 羽田野 英一

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 丸川 勝美

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(74)代理人 弁理士 小川 勝男

最終頁に続く

(54)【発明の名称】 伝票文書情報システム

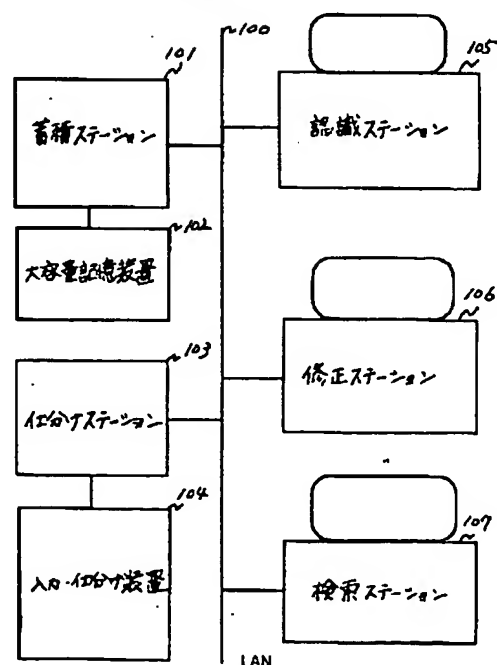
(57)【要約】

【目的】 複数の伝票文書の様式の種類を識別し、読み取り領域を自動的に設定し文字認識を行い、記載内容の妥当性を検定することにある。

【構成】 伝票文書情報システムの構成は、ネットワーク100に、仕分けステーション103、蓄積ステーション101、認識ステーション105、修正ステーション106、検索ステーション107が接続されている。

【効果】 複数の伝票文書を読み取ることができる。また、記載内容の妥当性を検定しており、伝票の記載内容を自動的に点検することができ、紙文書の仕分けを自動的に行なうことができ、仕分け作業を低減するという効果がある。

図1



## 【特許請求の範囲】

【請求項 1】診療報酬請求明細書の少なくとも画像入力、蓄積、検索、及び内容点検を行なう伝票文書情報システムにおいて、当該明細書の様式の種類を識別する手段と、当該様式の種類に応じて読み取り領域を抽出する手段と、読み取り領域を設定し文字認識を行なう手段と、当該認識結果のデータを基に記載内容の妥当性を検定する手段とを有することを特徴とする伝票文書情報システム。

【請求項 2】請求項 1 記載において、前記明細書の様式の種類を識別する手段は、明細書の種類をキー入力する手段と、様式の種類を予め辞書データとして格納する手段と、入力された明細書文書より様式データを抽出する手段と、当該辞書データと抽出した様式データとを照合する手段を有し、当該キー入力した明細書の種類により、辞書データとの照合を限定することを特徴とする伝票文書情報システム。

【請求項 3】紙文書の少なくとも画像入力、蓄積、及び検索を行なう伝票文書情報システムにおいて、伝票文書の種類を識別し文字認識を行なう手段と、認識結果によって整理番号を紙文書に打番する手段と、当該文書の種類に従って紙文書の仕分けを行なう手段とを有することを特徴とする伝票文書情報システム。

【請求項 4】紙文書の少なくとも画像入力、蓄積、及び検索を行なう伝票文書情報システムにおいて、伝票文書の種類を識別し文字認識を行なう手段と、当該認識結果に従ってバーコードの印字を紙文書の表面に行なう手段とを有することを特徴とする伝票文書情報システム。

【請求項 5】紙文書の少なくとも画像入力、蓄積、表示、及び検索を行なう文書情報システムにおいて、文書画像の複数領域を表示領域として設定する手段と、当該複数領域内において読み取り領域を抽出する手段と、当該表示領域の画像と認識結果を画面に同時に表示する手段と、認識結果を画面を介して修正するときに読み取り領域の画像を拡大表示する手段とを有することを特徴とする文書情報システム。

【請求項 6】紙文書の少なくとも画像入力、蓄積、検索、及び内容点検を行なう伝票文書情報システムにおいて、伝票文書の様式の種類を識別する手段と、当該様式の種類に応じて守秘内容の記載された領域を抽出する手段と、当該守秘領域の画像をマスクして表示する手段と、使用者に付与されたパスワードのレベルにより当該マスクの解除、及び設定を行なう手段とを有することを特徴とする伝票文書情報システム。

【請求項 7】紙文書の画像入力を行なう伝票文書情報システムにおいて、入力された伝票文書画像から枠線を抽出する手段と、標準的な伝票文書の枠線を登録する手段と、抽出した枠線と類似する枠線を登録データから選択する手段と、当該選択した登録枠線と抽出枠線を画面に同時に表示する手段とを有し、伝票文書を登録すること

を特徴とする伝票文書情報システム。

【請求項 8】紙文書の少なくとも画像入力、蓄積、検索、及び内容点検を行なう伝票文書情報システムにおいて、伝票文書の読み取り領域を抽出する手段と、読み取り領域内において文字認識を行なう手段と、認識結果の数値データが予め設定された範囲内にあるかどうかを判定する手段と、当該判定結果に従って伝票文書を蓄積する手段とを有し、当該範囲内にあると判定された伝票文書画像を選択し、当該画像のみを蓄積、及び検索することを特徴とする伝票文書情報システム。

【請求項 9】ネットワークを介して紙文書の仕分けを行なう文書情報システムにおいて、紙文書の表面画像を入力する入力ステーションと、入力した画像の種類を識別する認識ステーションと、識別結果に従って紙文書を区分する仕分けステーションと、画像を保管する蓄積ステーションを備え、ネットワークを経由して画像データ、当該種類データ、並びに仕分け制御情報を伝送することを特徴とする文書情報システム。

【請求項 10】紙文書の画像を入力し、当該紙文書画像を認識して種類を識別する仕分けステーションであって、紙文書を設定するホッパと仕分け後の紙文書を格納するスタッカを有し、スタッカが一杯になると警告音を発生し仕分け処理を一時停止することを特徴とする仕分けステーション。

【請求項 11】文書の少なくとも画像入力、蓄積、検索、及び内容点検を行なう文書情報システムにおいて、文書画像とその整理番号を格納する手段と、文書の整理番号を指定することによって整理用バーコードを求める手段と、紙文書の表面のバーコードを読み取る手段と、指定されたバーコードを付与された文書を選択する手段とを有することを特徴とする文書情報システム。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は様式や書式の異なる複数種類の伝票（例えば、診療報酬請求明細書等）の読み取り、仕分け、保管、検索などを行なう伝票文書情報システムに関する。

## 【0002】

【従来の技術】文書の様式は、記載項目及び項目の順番を規定するが、その絶対的な位置座標、寸法を規定しない。文書の書式は、記載項目及び項目の順番を規定するだけでなく、その絶対的な位置座標、寸法を規定している。従来の光学的文字認識装置では、文書の書式をフォーマットデータとして予め設定しており、文書の各項目の絶対的な位置座標、寸法が規定されている紙文書を読み取ることができた。しかしながら、様式のみが規定されている文書、例えば、法令文書などを読み取ることができなかった。

【0003】従来の伝票を読み取る文書情報システムは、特開昭 56-11573 号公報（特願昭 54-87

338号 昭和54(1979)年7月10日出願、発明者：香川正幸、出願人：富士通株式会社)に述べられているように、伝票に予め入力すべき項目を指示するマーク記入領域を設け、当該領域に記入されたマークを読み取ることによつて、伝票上の所定領域を読み取る方式が提案されている。しかしながら、従来の方式では、マーク記入領域を設けているため当該記入領域の紙面上の占める面積が多くなり、入力項目を増やすことができないという問題があった。また、マーク記入領域を新たに用意する必要があるため、既存の各種伝票をそのまま読み取ることができないという問題がある。さらに、様式が法令等で決められている伝票文書では、従来のようにマーク記入領域を設けることは法令上様式が規定されているため、認められず、マーク記入領域のない当該様式文書を読み取ることができないという問題があった。

【0004】また、従来の文書画像のファイル方式は、特開昭59-14070号公報(特願昭57-123460号 昭和57(1982)年7月15日出願、発明者：江尻公一、出願人：株式会社リコー)に述べられているように、中間調画像である写真や2値図形画像である棒グラフを含む文書の領域から特徴パラメータを抽出し、領域の画像と関連付けて記憶する画像ファイル方式が提案されている。しかしながら、従来の方式では、写真と棒グラフの識別を行なうことが述べられているだけであり、伝票のように野線並びに野線からなる枠と文字列からなる文書を識別することは述べられていない。また、伝票の文字列を文字認識することや、文書の領域を守秘の程度によってマスクするファイル方式について考慮されていない。

【0005】また、定形的な入力文書に対して文書の種類を自動的に識別し、すでに格納された文書と同一の分類コードを与えるファイリング方式が、特開昭61-75477号公報(特願昭59-196616号、昭和59(1984)年9月21日出願、発明者：中野康明、藤沢浩道、藤縄雅章、出願人：株式会社日立製作所)に述べられている。しかしながら、この従来方式では、画像を格納することが主であり、読み取り領域を設定し文字認識を行なうことや、当該認識結果のデータを基に記載内容の妥当性を検定することが考慮されていない。このため、文書の種類に応じて記載内容を読み取りかつ内容の妥当性を検定することができない。

【0006】

【発明が解決しようとする課題】上記各従来技術は、入力した紙文書そのものに対して整理番号若しくはバーコードを付与することが記載されておらず、蓄積した画像ファイルから所望の文書画像を表示、印刷することはできるが、原本である紙文書そのものを探索することは困難であり、原本を探す作業は多くの人手が必要な作業であった。また、紙文書そのものを仕分けすることが考慮されておらず、原本の分類、整理、収納に多くの人手を

要していた。

【0007】また、上記各従来技術では、蓄積した文書画像を表示、印刷する場合、守秘項目について考慮されておらず、例えば、文書にプライベートな内容や、ビジネス上重要な情報が記載されている場合でも検索結果の画像を画面にそのまま表示するか、若しくは、画像単位ごとに全面的表示を禁止するかのいずれかであり、文書の部分領域を守秘レベルに応じて見ることができなかった。

【0008】また、上記各従来技術では、入力した紙文書はすべて蓄積することが前提であり、内容によって、取捨選択して蓄積することは考慮されていない。このため、大量の文書を蓄積することとなり、小規模のファイルシステムでは記憶装置の容量を超えるという問題があった。

【0009】本発明の第1の目的は複数の伝票文書の様式の種類を識別し、読み取り領域を自動的に設定し文字認識を行い、記載内容の妥当性を検定することにある。

【0010】本発明の第2の目的は認識結果によって整理番号若しくはバーコードを紙文書に印刷し、紙文書の仕分けを行なうことにある。

【0011】本発明の第3の目的は文書画像から守秘内容の記載された領域を抽出し、守秘領域の画像をマスクして検索結果の表示、印刷を行なうことにある。

【0012】本発明の第4の目的は文書の記載内容によって蓄積の必要性の有無を自動的に判定し、記憶領域を節約することにある。

【0013】本発明の第5の目的はネットワークを介して認識、仕分け、蓄積を行なう文書情報システムを提供することにある。

【0014】

【課題を解決するための手段】上記第1の目的を達成するために、文書画像から様式の種類を識別して記載内容を読み取り、読み取り結果が記載項目ごとに登録している条件を満足するかどうかを判定するようにしたものである。

【0015】また、上記第2の目的を達成するためには、認識結果により仕分け機構並びに印刷機構を制御するようにしたものである。

【0016】また、上記第3の目的を達成するためには、文書画像から項目の記載された部分領域を自動的に抽出し、守秘領域に対してアクセスのためのパスワードを付与するようにしたものである。

【0017】上記第4の目的を達成するためには、記載内容例えば点数を読み取り、点数が一定値より大きい文書のみ、その表面画像を蓄積するようにしたものである。

【0018】上記第5の目的を達成するため、ネットワークに画像を圧縮して伝送するとともに、認識結果を伝送し、そのデータによってネットワークに接続された仕

分けステーション、蓄積ステーションを制御するようにしたものである。

#### 【0019】

【作用】複数種類の伝票に対してその表面画像を入力し、伝票文書の様式の種類を識別する。そして、様式の種類に応じて読み取り領域を抽出し、文字認識を行なう。認識結果のデータを基に記載内容の妥当性を予め設定している判定条件に従い検定する。さらに、認識結果によって整理番号を紙文書に打番若しくはバーコードを印字し、文書の種類に従って紙文書の仕分けを行なう。また、文書の種類に応じて守秘内容の記載された領域を抽出し、守秘領域の画像をマスクして表示できるようにする。

#### 【0020】

【実施例】以下、本発明を図面を用いて詳細に説明する。図1は本発明の一実施例である伝票情報システムの構成を説明する図である。

【0021】ネットワーク100には、入力・仕分けステーション103、蓄積ステーション101、認識ステーション105、修正ステーション106、検索ステーション107が接続されている。入力・仕分けステーション103は計算機より構成されている。当該ステーション103には入力・仕分け伝票文書の画像入力と伝票文書の仕分けを行なう入力・仕分け装置104が接続されている。当該装置には、文書を設定するホッパ、仕分け後の文書を格納するスタック及びその制御部があり、さらに、画像を入力するための光学的撮像素子、照明装置等の光学系が具備されている。ここでは、入力した画像をデータ圧縮し、当該ステーション103を介してネットワーク100に画像データを送出する。また、ネットワーク100から認識結果を受け取り、これを仕分け制御信号として文書の仕分けを行なう。蓄積ステーション101は計算機から構成され、光ディスクもしくはアレディスク等からなる大容量記憶装置102が接続されており、認識結果のデータ並びに入力された画像の蓄積を行なう。蓄積ステーション101では、認識結果に基づいて蓄積すべき文書を選別することができる。また、守秘内容の記載されている部分領域に対して、アクセス情報を付与して当該記憶装置102に蓄積する。認識ステーション105では、ネットワーク100を介して送られてきた画像に対して、伝票文書の種類を求めるシート識別処理を行ない、項目に記載されている文字列の認識を行なう。修正ステーション106では、認識結果の修正を画面を用いて対話的に行なう。ここでは、読み取り領域を拡大して画面に表示し、修正を容易にしている。検索ステーション107では、蓄積ステーション101に蓄積された伝票画像をネットワークを介して検索することができ、項目に記載されている内容を手掛かりに所望の伝票を探索し、画面に表示することができる。この時、守秘内容の記載されている部分領域に対し

てアクセス情報が付与されており、使用者のアクセスのレベルにより当該守秘項目をマスクして表示することができる。

【0022】図2は伝票文書の処理フローを説明する図である。伝票文書の一例として、診療報酬請求明細書（レセプト）の処理フローを示す。レセプトに記載された主要項目についてデータを入力し、計算機により資格記録の確認処理を行なう。本処理フローでは、先ず、ステップ200においてレセプトを入力用のホッパーに設定する。次いで、ステップ201でレセプトの仕分け処理を行なう。ここでは、レセプトの種類として、外来分レセプト210と入院分レセプト211にその様式を識別して仕分けを行なう。ステップ202では、初期情報を入力する。ここでは、初期情報として、レセプトの処理を各県単位に行なうための地域属性として、自県と他県の区別を入力する。また、レセプトの処理を行なう事務所名、診療区分、整理番号の初期値を登録する。なお、入院分レセプト211については、整理番号をステップ208で打番する。次いで、ステップ203、ステップ204で読み取りを実行する。ステップ203は、伝票の一括処理を行なうバッチ読み取り処理であり、認識結果の正常分ならびに不読（リジェクト）分を別のスタッカーに格納するとともに、整理番号を印字する。ステップ204は、逐次修正読み取りの処理であり、認識の結果リジェクトされたレセプトに対して、修正画面を介して対話的に認識結果を修正する。ステップ203で認識された正常分のデータと、ステップ204で修正され正常となったデータはフロッピーディスク等に格納する。また、ステップ203でリジェクトとなったレセプトは、別のスタッカーに格納されることになる。なお、入院分レセプト並びにステップ203のリジェクト分をまとめてステップ209でパンチ作業を行なう。レセプトに記載の診療内容をもとに、ステップ207に示すように資格記録確認処理等の機械処理を行う。ここでは、高額な点数が記載されているレセプトのみを選択してステップ205で光ディスクに格納している。ステップ207の資格記録の確認処理では、ステップ205で光ディスクに格納されたレセプトの検索206を行ない、所望のレセプトを画面に表示して資格記録の確認を行なう。

【0023】図3は伝票文書の画像入力、蓄積、検索、内容点検を行なうシステムの処理フローである。ステップ300で伝票文書の表面画像を入力し、ステップ301で伝票の様式の種類を識別するシート識別処理を行なう。当該シート識別の詳細処理フローは図4にて後述する。次いで、ステップ302において、当該様式の種類に応じて読み取り領域を抽出し、読み取り領域を設定する。当該読み取り領域設定ステップ302の詳細処理フローは図9にて後述する。ステップ303では読み取り領域内において文字認識を行ない、ステップ304で当

該認識結果のデータを基に記載内容の妥当性を検定する。当該ステップ303の詳細処理フローは図10にて後述する。

【0024】図4は、ステップ301のシート識別処理の詳細処理フローを示す図である。ここでは、入力した画像から枠及び文字行を抽出し、予め標準的な文書から登録したシート識別用辞書430とのマッチングを行なう。まず、ステップ400では入力画像から微小なノイズを除去する。次いで、ステップ401で黒画素の連続する線分であるランを用い罫線を抽出し、ステップ402で当該罫線の傾き角を算出する。ステップ403で画像を当該傾き角だけ回転させ、回転修正を行なう。そして再び、ステップ404で罫線を抽出する。図6は、この一連の処理により画像から水平罫線および垂直罫線を取り出し、その始点座標、終点座標を算出する過程の説明図である。伝票画像600に対して、601は水平罫線602、603、604を抽出した結果である。また、610は垂直罫線611から616を抽出した結果である。ステップ405では文字のつながりである文字行を長方形の形で抽出し、その角の座標を求める。文書の水平並びに垂直方向の位置ずれを修正するため、ステップ406で文書の基準点を求める。ステップ407では、水平罫線ならびに垂直罫線から枠の角に相当する交点の座標を算出する。各罫線は直線式で表現できるため、簡単に直線の交点座標を求めることができる。ステップ408では、求めた交点座標に対して、大きさの正規化を行なう。さらに、ステップ409では、交点を基に枠を生成する。図7は水平罫線及び垂直罫線から交点を求め、次いで枠を生成する過程を説明する図である。水平罫線602～604、並びに垂直罫線611～616をもとにその交点710～723を算出する。そして、交点を順次辿り枠701、702～706を生成する。図4のステップ410では、枠のマッチングを行なう。この時、シート識別用辞書430に格納されている複数個の標準の枠とのマッチングを行なう。当該辞書430には、枠位置及び文字行の位置が予め登録されており、その登録のフローは、まず、ステップ420で標準の伝票画像を入力し、次いで、ステップ421で入力すべき枠及び文字行の位置を画面から指定する。ステップ422では位置並びに回転の補正を行ない、辞書として登録する。さらに、ステップ411では、文字行マッチングを行なう。この時、シート識別用辞書430に格納されている複数個の標準の文字行とのマッチングを行なう。文字行マッチングの詳細は図8において後述する。上に述べた枠マッチング410と文字行マッチング411によって求めた結果に従って、標準の伝票の内、もっとも類似している伝票が有するシート種類番号をステップ412で決定する。さらに、ステップ413で、生成した枠位置ならびに文字行の位置を出力する。また、ステップ414でシート種番号を出力する。

【0025】図5は位置ずれ・回転・伸縮を有する認識座標系を示している。伝票シート500に506で示す枠罫線が存在している。認識座標系を伸縮の無い認識座標系 $(X_r, Y_r)$ 511で、伸縮のある認識座標系 $(X_m, Y_m)$ 512で表現する。これらの座標系は、センサと光学系からなる観測座標系である。また、フォーマットデータ生成用座標系 $(X_f, Y_f)$ 513はシートの端を基準にした座標系であり、実物の寸法を表す座標系である。伸縮の無い認識座標系で表現したシート内の目標点 $(x_r, y_r)$ 507は、認識座標系の原点510とフォーマットデータ生成用座標系の原点509とのずれ量 $(dx, dy)$ と回転角 $\theta$ により表現できる。

【0026】

$$\begin{aligned} x_r &= dx + x_f \cdot \cos \theta + y_f \cdot \sin \theta \\ y_r &= dy - x_f \cdot \sin \theta + y_f \cdot \cos \theta \end{aligned}$$

ここで、 $(x_f, y_f)$ はフォーマットデータ生成用座標系で表した目標点507の座標である。さらに、伸縮のある認識座標系で表現したシート内の目標点507

$(x_m, y_m)$ は、伸縮率 $m_x, m_y$ を、

【0027】

$$\begin{aligned} \text{【数2】 } x_m &= m_x \cdot x_r \\ y_m &= m_y \cdot y_r \end{aligned}$$

とすると、

【0028】

【数3】

$$\begin{aligned} x_m &= m(dx + x_f \cdot \cos \theta + y_f \cdot \sin \theta) \\ y_m &= m(dy - x_f \cdot \sin \theta + y_f \cdot \cos \theta) \end{aligned}$$
と表すことができる。シート内の読み取りでは、上式に従って、フォーマットデータを変換して画像にアクセスし文字認識を行なう。また、位置合わせ用基準点508の認識座標系の値を伸縮の無い認識座標系 $(x_{r0}, y_{r0})$ 、伸縮のある認識座標系 $(x_{m0}, y_{m0})$ で表す。また、フォーマットデータ生成用座標系の値を $(x_{f0}, y_{f0})$ で表す。位置ずれ量 $(dx, dy)$ は、

【0029】

$$\begin{aligned} \text{【数4】 } dx &= (1/m_x) \cdot x_{m0} - x_{f0} \cdot \cos \theta - y_{f0} \cdot \sin \theta \\ dy &= (1/m_y) \cdot y_{m0} + x_{f0} \cdot \sin \theta - y_{f0} \cdot \cos \theta \end{aligned}$$

で表すことができる。回転角 $\theta$ は罫線の傾きを基に検出する。次いで、伸縮率 $m_x, m_y$ を求めるが、これは、例えば、一番外の黒線枠の横幅、縦幅を検出し、フォーマットデータの該当する横幅、縦幅と比較して算出する。さらに、位置合わせ用基準点の認識座標系における位置 $(x_{m0}, y_{m0})$ を求め、フォーマットデータに予め設定している当該点の位置 $(x_{f0}, y_{f0})$ (フォーマットデータ生成用座標系で表現)と比較して、位置ずれ量 $(dx, dy)$ を求める。シート識別では、伸縮のある認識座標系において検出した目標点の座標 $(x$



m, y m)をフォーマットデータ生成用座標系(x f, y f)に変換し、これをシート識別用辞書の書式の座標データと比較する。

【0030】文字行の相対的な位置関係を表現する方式を図29に示す。今、注目する文字行2900に対して、8方向の位置関係を表現している。方向を東西南北で表すと、注目する文字行2900の北方向2901には文字行2911がある。また、南方向2905には文字行2915が、東方向2903には文字行2913が、西方向2907には文字行2917が、それぞれある。さらに、注目する文字行2900の北東方向2902には文字行2912が、北西方向2908には文字行2918が、南東方向2904には文字行2914が、南西方向2906には文字行2916が、それぞれある。このように、注目する文字行に対して、8方向の位置関係を有する文字行を表現する。

【0031】図8はステップ411の文字行マッチングの処理方を説明する図である。伝票文書の文字行の部分領域の例を810、811、812に示す。文字行領域810では、抽出した文字行を801、802、803示しているが、これらの相対的な位置関係は上下の南北関係になっている。いま、注目する文字行を801とすると、文字行802は文字行801に対して南方向820にある。また、文字行802に対して、文字行803は、同じく、南方向821にある。一方、文字行領域811の例では、文字行804、805、806の相対的な位置関係は左右の東西関係である。文字行804に対して、文字行805は東方向822にあり、文字行805に対して806も又、東方向823にある。さらに、文字行領域812の例では、斜め方向の位置関係にあり、文字行807に対して、文字行808は北東方向824にある。また、文字行808に対して文字行809は、同じく、北東方向825にある。ステップ411の文字行マッチングでは、このような文字行の相対的な位置関係が予めシート識別用辞書430には登録されており、入力画像から抽出した文字行の相対的な位置関係と合致するかどうかを判定する。

【0032】図9は図3のステップ302で示した読み取り領域設定処理の詳細処理フローである。先のシート識別のステップ301で得られたシート種番号をステップ900で入力する。次いで、ステップ901で読み取り領域ごとの処理に移り、ステップ902でシート種類ごとに登録されているフォーマットパラメータ910から、該当する読み取りのフォーマットパラメータを読み込む。ここで、フォーマットパラメータとしては、文字種、文字数、読み取りフィールドの位置が具備されている。ステップ903では、フォーマットパラメータから読み込んだ読み取りフィールド位置とステップ406で求めた基準点位置からの位置ずれ量を基に、読み取り領域を算出する。

【0033】図10はステップ303で示した読み取り領域内文字認識処理の詳細処理フローである。ステップ1000で読み取りフィールドごとに、ステップ1001から1004までの処理を行なう。ステップ1001では読み取り領域内の文字行を抽出する。ステップ1001では、予め、読み取り領域内に印刷されている不要文字を除去する。そして、ステップ1003で抽出した文字行を一文字単位に分離する。ステップ1004では、一文字ごとの文字認識を行なう。この文字認識は、従来技術を用いて容易に実現できる。

【0034】図11は伝票文書を読み取り、当該読み取り結果をもとに伝票文書を選択して蓄積する伝票文書情報システムの処理フローを示す図である。ステップ1100で文書がつきるまで、ステップ1101以下の処理を繰り返す。ステップ1101で伝票文書の画像入力を行なう。次いで、ステップ1102で、ステップ301と同様のシート識別処理を行なう。次いで、ステップ1103で同じく、読み取り領域を抽出し、ステップ1104で読み取り領域内において文字認識を行なう。ステップ1105で、認識結果の数値データが予め設定された範囲内にあるかどうかを判定し、もし設定された範囲内にある場合は、ステップ1106で光ディスク等の記憶装置に当該画像を蓄積する。もし、設定された範囲内にはない場合は、記憶装置に当該画像を蓄積しない。上述の一連の処理をステップ1100で文書が尽きるまで行なう。この処理により入力した紙文書はすべて蓄積する必要はなく、記載内容によって、取捨選択して蓄積することができる。このため、小規模のファイルシステムでは記憶の容量を節約できるという効果がある。

【0035】伝票文書の画像入力、蓄積、検索、内容点検を行なう文書情報システムにおいては、使用者のアクセスレベルに応じて守秘領域の画像をマスクして表示する。図12は伝票文書画像から守秘領域を登録する処理フローを示す。ステップ1200で画像を入力し、ステップ1201でシート識別処理301を行なう。そして、ステップ1202で守秘領域を抽出する。ここでは、先のシート識別のステップ1201でシート種と抽出した領域の位置を出力しており、領域ごとに守秘領域であるかどうかを判定する。この時、守秘領域であるかどうかは、予め登録しているシート識別用辞書430を読み取るものとする。当該辞書430には、枠で示された領域ごとに守秘のレベルが具備されている。ステップ1203で守秘領域の座標を画像データとともに記憶装置に格納する。

【0036】図13は伝票画像に対して、使用者のアクセスレベルにより守秘領域の画像をマスクして表示する処理のフローである。まず、ステップ1300で使用者のパスワードを入力する。そして、ステップ1301で使用者のアクセスレベルを図16に示す使用者アクセスレベルテーブル1600を用いて設定する。図16は使

用者アクセステーブルの説明図である。使用者名1601、パスワード1602に対して、アクセスレベル1603が格納されており、当該使用者に対応するアクセスレベルを設定する。次いで、ステップ1302で領域がつきるまで、ステップ1303、1304の処理を繰り返す。ステップ1303では、領域ごとに設定されているアクセスレベルを領域アクセステーブル1500から読み出す。図15は領域アクセステーブルを示す。当該テーブル1500は、標準の伝票文書に対して、領域に付与された識別番号1501ごとにアクセスレベル1502が付与されている。そして、ステップ1304で使用者のアクセスレベルが領域のアクセスレベルより強いかどうかの判定を行なう。もし、使用者のアクセスレベルが強い場合は、ステップ1305で当該領域のマスクを解除する。一方、その反対に、領域のアクセスレベルの方が使用者のアクセスレベルより強い場合は、ステップ1306で当該領域にマスクを設定する。以上の一連の処理をステップ1302で領域がつきるまで繰返し行なう。そして、ステップ1307で伝票画像を表示する。ここでは、領域ごとに表示処理を実行し、領域にマスクが設定されている部分画像はそのまま表示せず、例えば、黒画像により塗りつぶした画像を表示する。領域のマスクが解除されている部分画像はそのまま表示する。本実施例では、蓄積した文書画像を表示、印刷する場合、守秘項目についてアクセスレベルが付与されており、例えば、文書にプライベートな内容や、ビジネス上重要な情報が記載されている場合に、文書の部分領域を守秘レベルに応じて見ることができ、機密を保つことができるという効果がある。

【0037】図14は伝票文書画像の表示の一例である。表示画面1400において、伝票の各項目に対応した領域1401、1402、1403、1404、1405、1406、1407、1408、1409が設定される。この時、使用者のアクセスレベルに応じて、斜線部分1410、1411、1412、1413、1414において、守秘のためにマスクが設定されており、当該領域は黒く塗りつぶされることになる。

【0038】図17は伝票文書の読み取り結果によってバーコードを当該文書に印刷する処理フローを示す。ステップ1700で文書画像を入力する。入力した画像に対して、ステップ1701で伝票文書の種類を識別するシート識別を行ない、ステップ1702で文字認識によりシート内の領域の読み取りを行なう。そして、ステップ1703で当該認識結果に従ってバーコードの印字を紙文書の表面に行なう。

【0039】図18はバーコードにより文書を探索する処理フローを示す。ステップ1800で、探索したい文書の整理番号を入力すると、ステップ1801で整理番号に該当するバーコードを設定する。そして、ステップ1802で文書が尽きるまで、ステップ1803、18

04、1805の処理を繰り返す。まず、ステップ1803では、伝票文書の紙送りを実行し、紙表面に印刷されているバーコードをステップ1804で読み取る。次いで、ステップ1805において、読み取ったバーコードがステップ1801で設定されたバーコードと一致するかどうかを判定する。もし、一致すれば、所望の伝票文書であるとして、ステップ1806で選択スタックに格納する。もし、ステップ1805で一致しなければ、ステップ1807で通常のスタックに格納し、ステップ1802で文書がつきるまで処理を繰り返す。本実施例により原本にバーコードを印刷しているため、蓄積した画像ファイルから所望の文書画像を表示、印刷するだけでなく、原本である紙文書そのものをバーコードを用いて容易に探索することができるという効果がある。図19は伝票文書の種類を識別し仕分けを行なう処理フローである。ステップ1900で伝票文書画像を入力し、シート識別用辞書1910によってステップ1901でシート識別を行なう。次いで、ステップ1902でシート内読み取りを実行する。この時、予め登録しているフォーマットパラメータ1911と文字認識用辞書1912を用いる。ステップ1903で、シート識別の結果であるシートの種類によって伝票文書の仕分けを行なう。本実施例により、従来人手により行なわれていた伝票文書の仕分けを自動的に行なえるようになり、人手による作業量を低減することができる。

【0040】図20はネットワークに接続された入力・仕分けステーションと認識ステーションの動作制御を説明する図である。ネットワーク2000には入力・仕分けステーション2002と認識ステーション2001が接続されている。処理の制御は、まず、認識ステーション2001から全体システムの起動を入力・仕分けステーション2002に送る。入力・仕分けステーション2002、この起動指令を受け、ホッパに設定された伝票文書は入力装置2004のスキヤナで画像を入力する。画像は入力装置2004内にある画像プロセッサにおいてデータ圧縮され、ネットワーク2000を介して認識ステーション2001に転送される。転送された画像は伸張され、シート識別と読み取りを行なう。認識結果は、ネットワーク2000を介して入力・仕分けステーション2002に転送され、仕分け装置2005のスタックを選択する。上記の処理をホッパの設定されている伝票文書がなくなり、ホッパが空になるまで行なう。なお、スタックが途中で満杯になった場合は、警告音が発生し、一時停止する。ホッパが空になれば、入力・仕分けステーション2002から終了指令を認識ステーション2001に送信し、全体システムの処理が終了する。

【0041】図21は伝票文書の種類を識別し文字認識を行い認識結果によって整理番号を紙文書に打番して仕分けを行なう伝票文書処理のフローである。ステップ2100で伝票文書の画像を入力する。ステップ2101

で伝票の種類を識別するシート識別を実行する。ステップ2102でシート内読み取りを行ない、当該認識結果によってステップ2103で整理番号を打番する。ステップ2104では伝票の種類、若しくは、記載内容によって仕分けを行ない所定のスタッカに格納する。

【0042】図22は整理番号を打番した伝票文書の一例である。伝票文書2200において、2201で示す整理番号を打番している。文書の端部の位置を検出しているため、当該整理番号は予め設定している座標位置に打番することができる。また、打番のためのインク色を指定色に切り換えることができる。

【0043】図23は入力・仕分けの構成を示す図である。ホッパ2300に伝票文書2320が設定されている。2304は計算機であり入力・仕分け全体を制御するとともに、ネットワーク2321との通信制御を行なう。ホッパ制御部2310は、ホッパ機構部2315を制御し、順次、紙をセンサ部2313に送る。センサ制御部2311は画像信号をデジタル化し計算機2304に送る。2314は打番機構であり、打番制御部2312によって計算機2304から送られてきた整理番号を印字する。仕分け機構部2316、2317、2318は、計算機2304の指令により、送られてきた伝票文書を所定のスタッカに格納するよう制御されている。スタッカ2301、2302、2303には選択された伝票文書が順次格納されている。各スタッカには文書の満杯を検知するセンサ2307、2308、2309が具備されており、満杯検知センサ制御部2306によってスタッカ満杯信号が計算機2304に送られる。2305は警告音発生部であり、スタッカの満杯が検知されると、警告音を発生し、計算機2304は処理を一時停止する。

【0044】図24は伝票文書情報システムの画面の遷移図の一例である。システム立ち上げ時のメイン画面2401を図25に示す。メイン画面2500は、選択項目2501並びに選択項目の入力プロンプト2502を具備している。入力方法指定画面2402によって、2種類の入力を切り換えることができ、その一つはスキャナ入力画面2403、もう一つはファイル入力画面2404で行なう。2405は画像表示画面であり、指定した入力方法で画像を入力し、その入力画像を当該画面で表示確認することができる。当該画面から、メイン画面2401に戻ることができ、例えば、様式／書式識別画面2406に遷移する。ここでは、図26に示すように、抽出した枠2601を画面2600に表示することができる。また、画面2602に示すように、抽出した枠2603と辞書として具備していた枠2604を重ねあわせて表示することもできる。2408は認識結果画面であり、文書画像と認識結果を同時に表示し、修正を行なうことができる。図27は認識結果画面を説明する図である。伝票文書2700に対して、上部画像表示エ

リア2701と下部画像表示エリア2702を拡大して表示する。認識結果画面2710において、2711に上部画像を拡大して表示している。また、2712に下部画像を拡大して表示している。認識結果は2713の部分に表示しており、修正を対話的に行なうことができる。図28は、認識結果画面の他の実施例を説明する図である。本実施例では、認識結果画面を2画面として、第1画面2800には上部画像2801を表示し、認識結果は2802の部分に表示している。また、第2画面2803には、下部画像2804を表示しており、同じく認識結果を2805に表示し、対話的に修正を行なう。

【0045】図30はステップ304の妥当性検定の詳細処理フローであり、レセプトを対象とした例である。ステップ3000で、性別と請求内容の検定を行なう。性別によって医療行為が異なるものがあり、性別により検査や手術等の請求内容が該当するかどうかを検定する。さらに、ステップ3001では、年齢と請求内容の検定を行なう。ここでは、年齢により調剤や手術の内容に制限があり、請求内容が当該年齢において適切であるかの検定を行なう。さらに、ステップ3002では、年齢と伝票様式の検定を行なう。ここでは、年齢により、例えば、新生児、乳幼児、老人等により伝票の様式が予め規定されており、シート識別で求めた伝票の様式が記載されている年齢と合致するかどうかの検定を行なう。また、ステップ3003で保険証番号と登録されている氏名との一致を判定する。さらに、ステップ3004と3005において、診療実日数及び合計点の検定を行なう。

#### 【0046】

【発明の効果】文書画像から様式の種類を識別して記載内容を読み取り、読み取り結果が記載項目ごとに登録している条件を満足するかどうかを判定するようにしており、複数の伝票文書を読み取ることができる。また、読み取り領域を自動的に設定し文字認識を行い、記載内容の妥当性を検定しており、伝票の記載内容を自動的に点検することができる。認識結果により仕分け機構並びに印刷機構を制御するようにしており、認識結果によって整理番号若しくはバーコードを紙文書に印刷し、紙文書の仕分けを自動的に行なうことができ、仕分け作業を低減するという効果がある。守秘領域に対してアクセスのためのパスワードを付与するようにしており、守秘領域の画像をマスクして検索結果の表示、印刷を行なうことができ、システムの信頼性を向上させる効果がある。記載内容例えば点数を読み取り、点数が一定値より大きい文書のみ、その表面画像を蓄積するようにしており、記憶領域を節約できるという効果がある。

#### 【図面の簡単な説明】

【図1】伝票情報システムの構成を説明する図である。

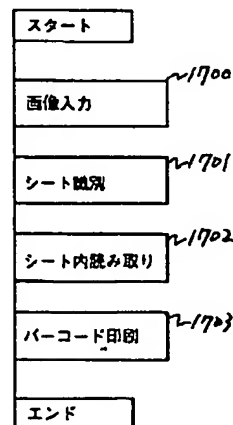
【図2】伝票文書の処理フローを説明する図である。

【図18】バーコードにより文書を探索する処理フローの説明図である。

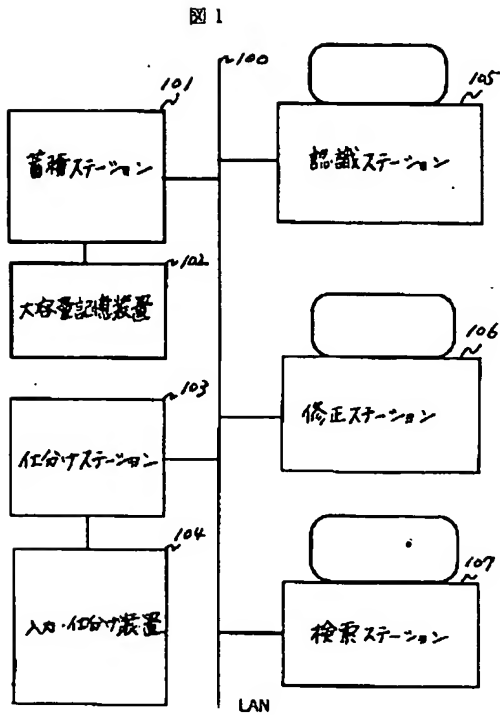
【図30】 妥当性検定の詳細処理フローである。

100…ネットワーク、101…蓄積ステーション、105…認識ステーション、103…入力・仕分けステーション、301…シート識別処理ステップ、304…妥当性検定ステップ、410…枠マッチングステップ、411…文字行マッチングステップ、910…フォーマットパラメータ、1306…領域のマスク設定、1703…バーコード印刷ステップ、1804…バーコード読み取り、2103…整理番号打番、2104…仕分け処理。

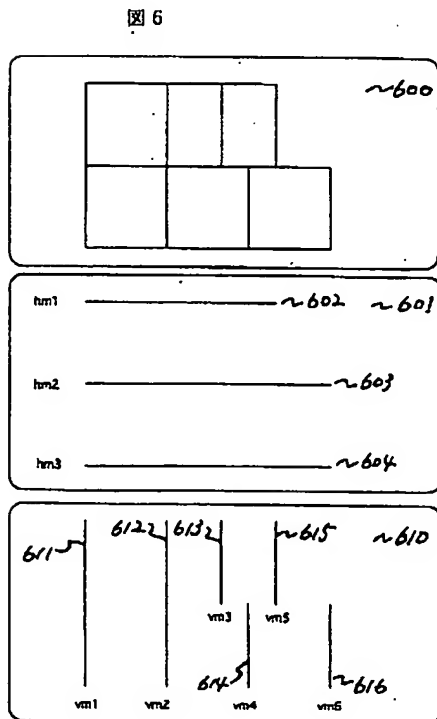
【图 17】



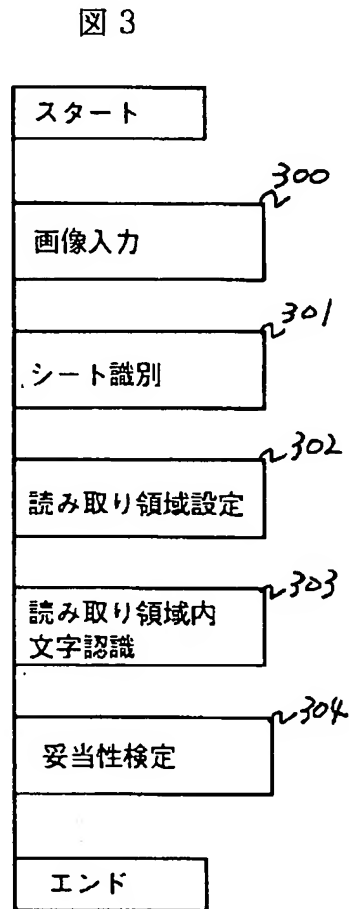
【図1】



【図6】

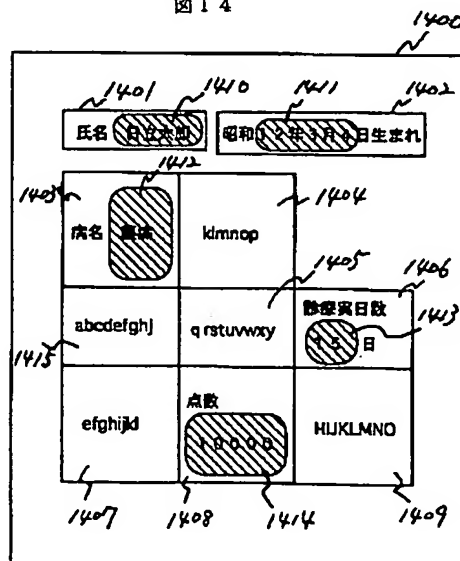


【図3】

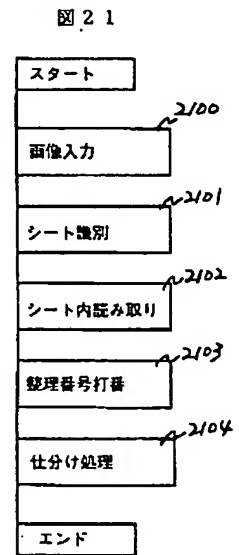


【図14】

図14



【図21】



【図2】

【図22】

図2

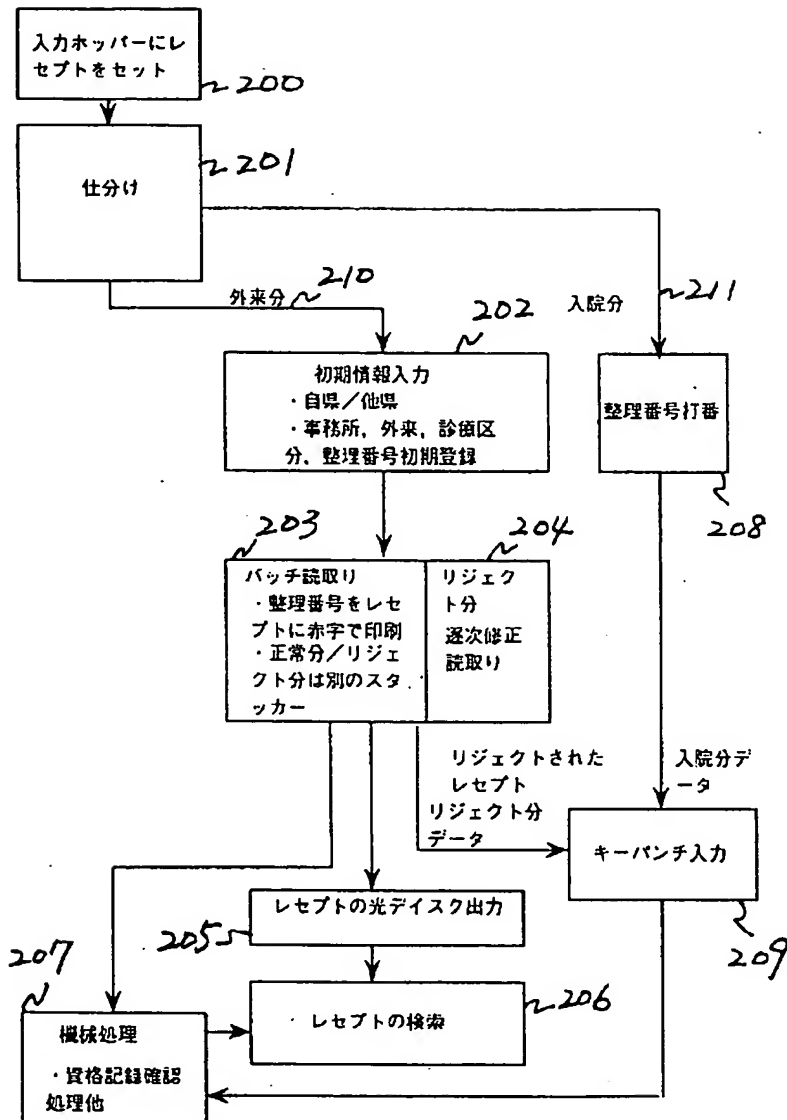


図22

図22は、医療記録入力用フォームのスクリーンショットを示す。フォームには、診療科目、年月日、患者情報などの入力欄があり、下部には「ナンバリング」のボタンが配置されている。

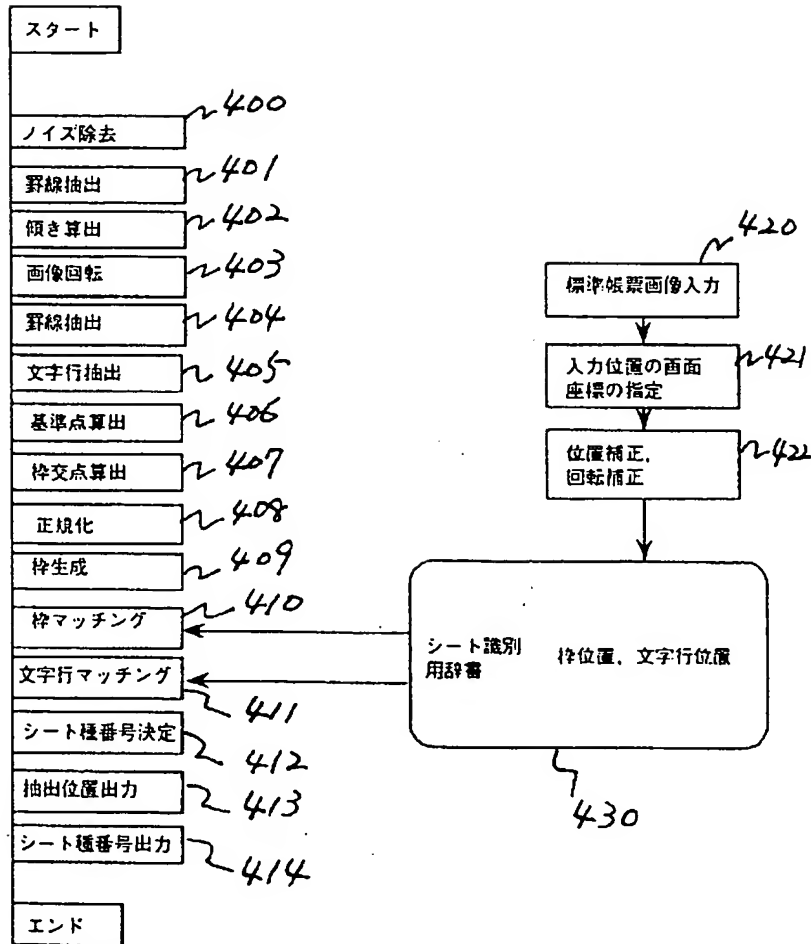
【図26】

図26

図26は、医療記録出力用フォームのスクリーンショットを示す。このフォームは、患者の診療履歴や検査結果を詳細に記録するための欄が複数あり、下部には「2601」から「2604」までの番号が振られている。

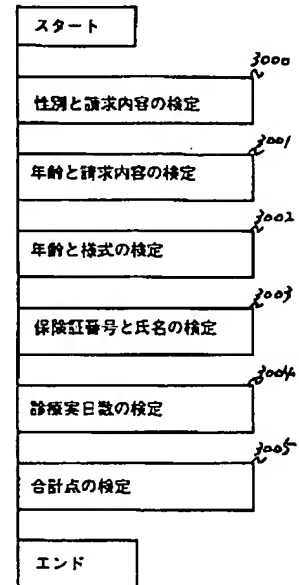
【図4】

図4



【図30】

図30



【図9】

図9

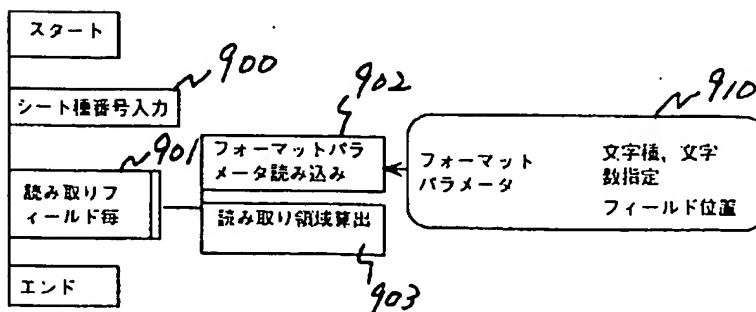
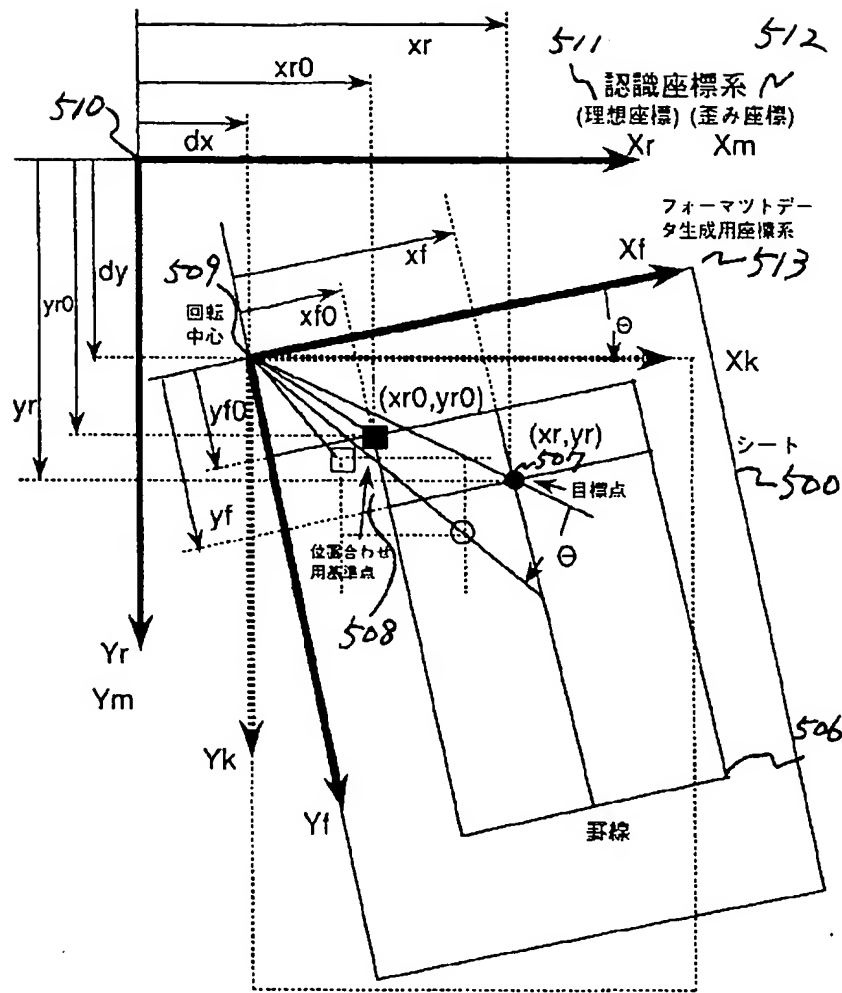


图 5



【图 16】

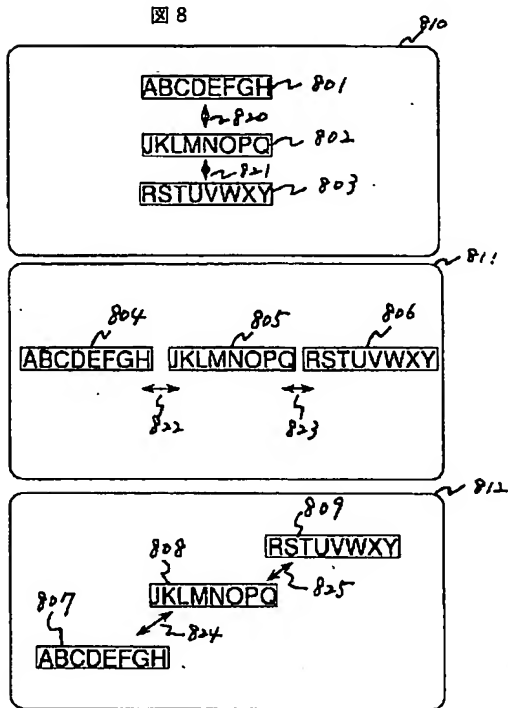
16

使用者名 1601 パスワード 1602 アクセスレベル 1600

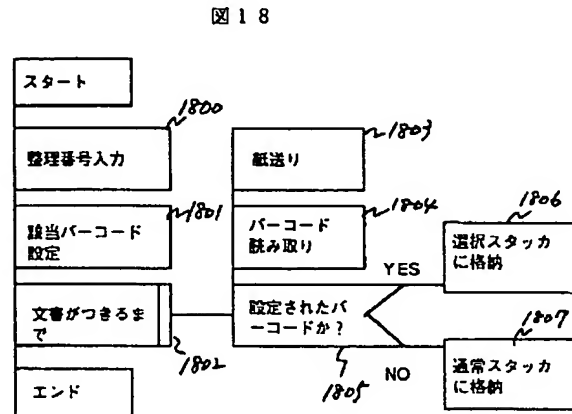
|           |     |   |
|-----------|-----|---|
| shima     | abc | 0 |
| marukawa  | bcd | 2 |
| koga      | cde | 3 |
| nakashima | def | 4 |



【図 8】

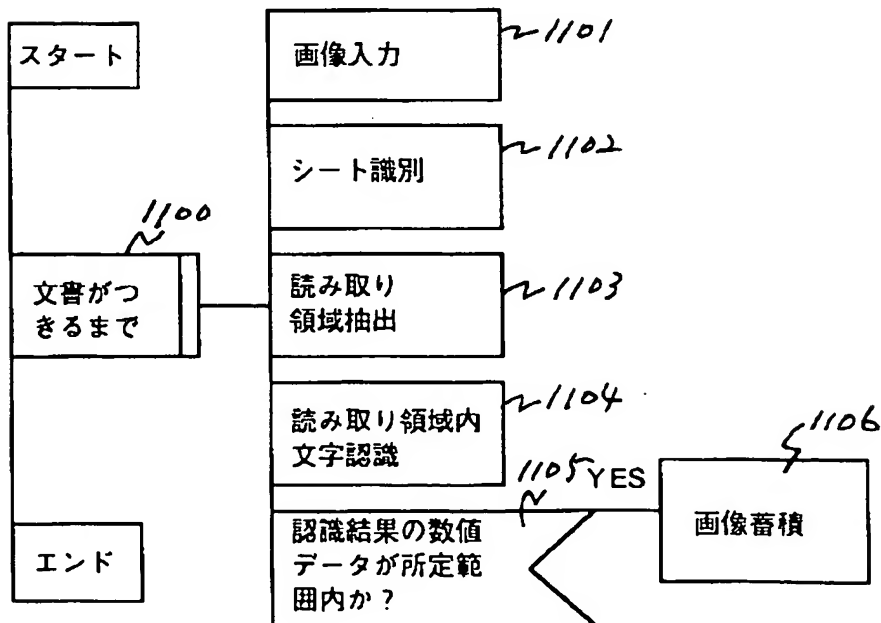


【図 18】



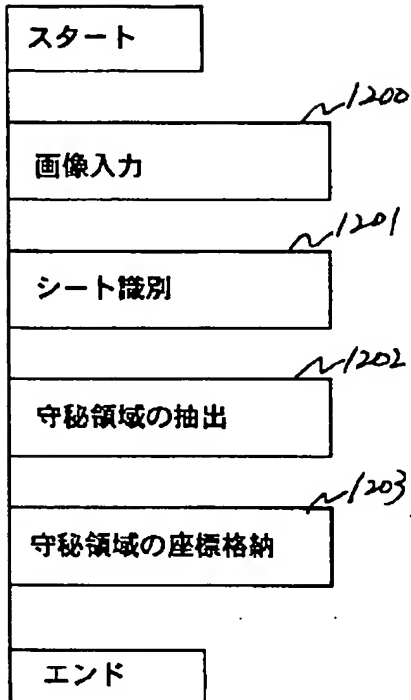
【図 11】

図 1 1



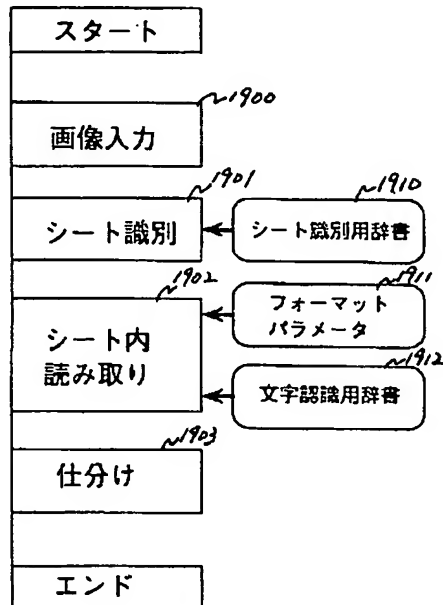
【図12】

図12



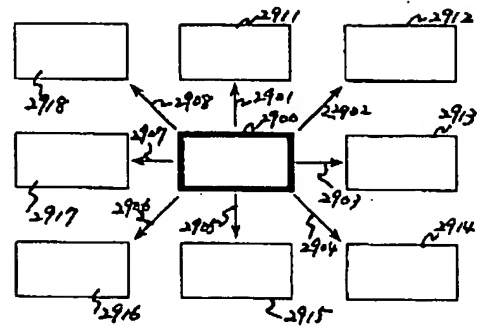
【図19】

図19



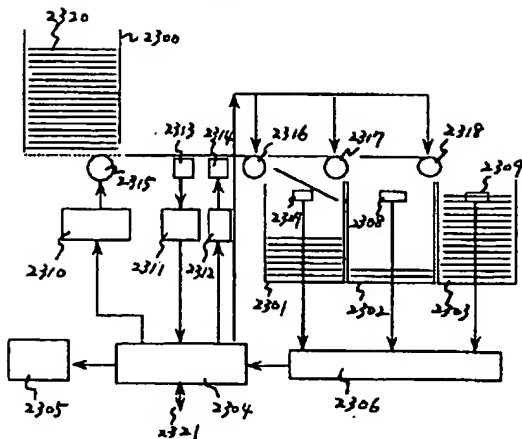
【図29】

図29



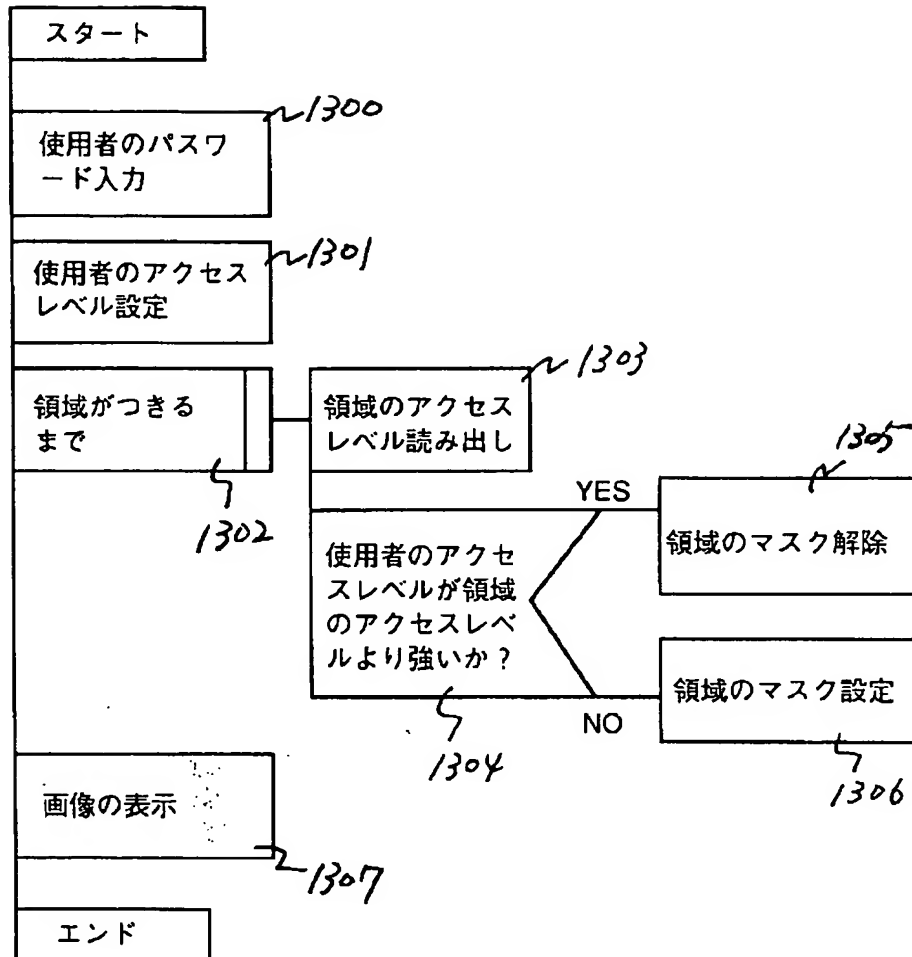
【図23】

図23



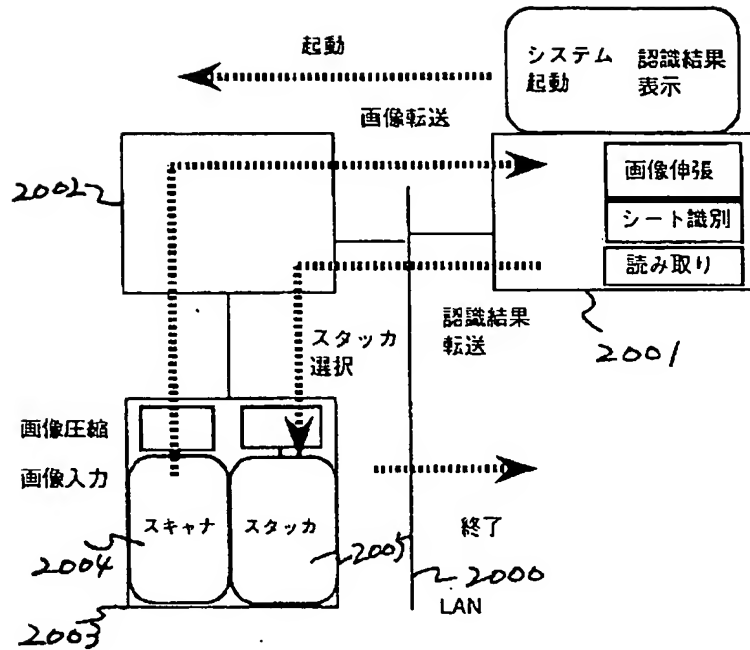
【図13】

図13



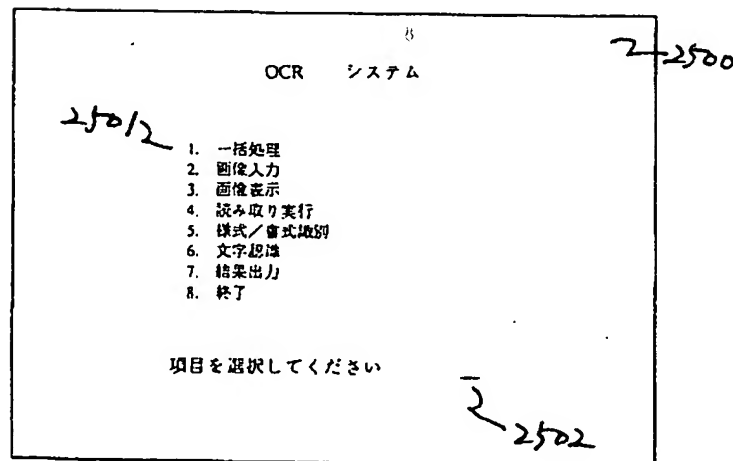
【図20】

図 20



【図25】

図 25





【図28】

図 28

2801 上部画像 2800

シーケンス番号 00123

登録情報画面 平成 5 年 12 月

氏名

2802

2804 下部画像 2803

シーケンス番号 00123

合計 決定

2805

|      |             |         |         |
|------|-------------|---------|---------|
| 登録年月 | 平成 5 年 12 月 | 登録情報コード | 0123456 |
| 性別   | 男           | 生年      | 5 年     |
| 登録日数 | 1 日         | 合計      | 123 日   |

フロントページの続き

(72) 発明者 古賀 昌史

東京都国分寺市東恋ヶ窪 1 丁目 280 番地  
株式会社日立製作所中央研究所内

(72) 発明者 中島 和樹

東京都国分寺市東恋ヶ窪 1 丁目 280 番地  
株式会社日立製作所中央研究所内

(72) 発明者 門田 彰三

神奈川県小田原市国府 2880 番地 株式会社  
日立製作所ストレージシステム事業部内

(72) 発明者 栗野 清道

神奈川県小田原市国府 2880 番地 株式会社  
日立製作所ストレージシステム事業部内

(72) 発明者 杉本 建行

神奈川県小田原市国府 2880 番地 株式会社  
日立製作所ストレージシステム事業部内

【公報種別】特許法第 17 条の 2 の規定による補正の掲載

【部門区分】第 6 部門第 3 区分

【発行日】平成 13 年 2 月 9 日 (2001. 2. 9)

【公開番号】特開平 7-114616

【公開日】平成 7 年 5 月 2 日 (1995. 5. 2)

【年通号数】公開特許公報 7-1147

【出願番号】特願平 5-262100

【国際特許分類第 7 版】

G06K 9/00

G06F 19/00

【F I】

G06K 9/00 K

G06F 15/22

【手続補正書】

【提出日】平成 12 年 3 月 28 日 (2000. 3. 28)

【手続補正 1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正内容】

【特許請求の範囲】

【請求項 1】伝票文書を画像として入力し、前記画像の蓄積、検索、及び内容点検を行なう伝票文書情報システムにおいて、前記伝票文書の様式の種類を識別する手段と、前記様式の種類に応じて読み取り領域を抽出する手段と、読み取り領域を設定し文字認識を行なう手段と、前記文字認識結果のデータを基に記載内容の妥当性を検定する手段とを有することを特徴とする伝票文書情報システム。

【請求項 2】請求項 1 記載の伝票文書情報システムにおいて、前記伝票文書の様式の種類を識別する手段は、前記伝票文書の種類をキー入力する手段と、様式の種類を予め辞書データとして格納する手段と、入力された明細書文書より様式データを抽出する手段と、前記辞書データと抽出した様式データとを照合する手段を有し、前記キー入力した前記伝票文書の種類により、辞書データとの照合を限定することを特徴とする伝票文書情報システム。

【請求項 3】伝票文書を画像として入力し、前記画像の蓄積、及び検索を行なう伝票文書情報システムにおいて、伝票文書の種類を識別し文字認識を行なう手段と、前記認識手段の認識結果によって整理番号を紙文書に打番する手段と、前記伝票文書の種類に従って紙文書の仕分けを行なう手段とを有することを特徴とする伝票文書情報システム。

【請求項 4】伝票文書を画像として入力し、前記画像の蓄積、及び検索を行なう伝票文書情報システムにおい

て、伝票文書の種類を識別し文字認識を行なう手段と、前記認識手段の結果に従ってバーコードの印字を紙文書の表面に行なう手段とを有することを特徴とする伝票文書情報システム。

【請求項 5】伝票文書を画像として入力し、前記画像の蓄積、表示、及び検索を行なう文書情報システムにおいて、伝票文書の複数領域を表示領域として設定する手段と、前記複数領域内において読み取り領域を抽出する手段と、伝票文書を認識する手段と、前記表示領域の画像と、前記認識手段の認識結果を画面に同時に表示する手段と、前記認識結果を画面を介して修正するときに読み取り領域の画像を拡大表示する手段とを有することを特徴とする文書情報システム。

【請求項 6】伝票文書を画像として入力し、前記画像の蓄積、検索、及び内容点検を行なう伝票文書情報システムにおいて、伝票文書の様式の種類を識別する手段と、前記様式の種類に応じて守秘内容の記載された領域を抽出する手段と、前記守秘領域の画像をマスクして表示する手段と、使用者に付与されたパスワードのレベルにより前記マスクの解除、及び設定を行なう手段とを有することを特徴とする伝票文書情報システム。

【請求項 7】伝票文書を画像として入力する伝票文書情報システムにおいて、入力された伝票文書画像から枠線を抽出する手段と、標準的な伝票文書の枠線を登録する手段と、抽出した枠線と類似する枠線を登録データから選択する手段と、前記選択した登録枠線と抽出枠線を画面に同時に表示する手段とを有し、伝票文書を登録することを特徴とする伝票文書情報システム。

【請求項 8】伝票文書を画像として入力し、前記画像の蓄積、検索、及び内容点検を行なう伝票文書情報システムにおいて、伝票文書の読み取り領域を抽出する手段と、読み取り領域内において文字認識を行なう手段と、前記認識手段の認識結果の数値データが予め設定された範囲内にあるかどうかを判定する手段と、当該判定結果

に従って伝票文書を蓄積する手段とを有し、当該範囲内にあると判定された伝票文書画像を選択し、前記選択された画像のみを蓄積又は検索することを特徴とする伝票文書情報システム。

【請求項 9】ネットワークを介して伝票文書の仕分けを行なう文書情報システムにおいて、伝票文書の表面画像を入力する入力ステーションと、入力した画像の種類を識別する認識ステーションと、前記認識ステーションの識別結果に従って紙文書を区分する仕分けステーションと、画像を保管する蓄積ステーションを備え、前記ネットワークを経由して画像データ、前記種類に関するデータ、並びに仕分け制御情報を伝送することを特徴とする文書情報システム。

【請求項 10】伝票文書の画像を入力し、前記伝票文書

画像を認識して種類を識別する仕分けステーションであって、紙文書を設定するホッパと仕分け後の紙文書を格納するスタッカを有し、スタッカが一杯になると警告音が発生し仕分け処理を一時停止することを特徴とする仕分けステーション。

【請求項 11】伝票文書を画像として入力、前記画像の蓄積、検索、及び内容点検を行なう文書情報システムにおいて、文書画像とその整理番号を格納する手段と、前記整理番号を指定することによって整理用バーコードを求める手段と、伝票文書の表面のバーコードを読み取る手段と、指定されたバーコードが付与された文書を選択する手段とを有することを特徴とする文書情報システム。